THE PRIMARY STRUCTURE OF HISTONES H2B FROM SPERM OF THE SEA URCHINS PARECHINUS ANGULOSUS AND PSAMMECHINUS MILIARIS

by

Marie Shields Strickland

Submitted in fulfilment of the requirements

for the degree of

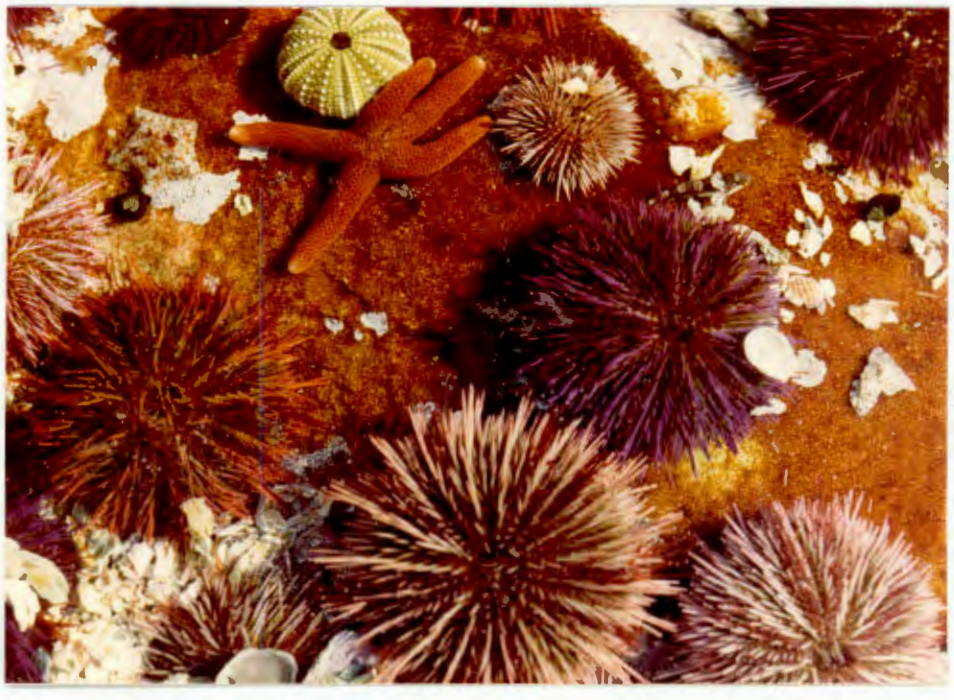
Doctor of Philosophy,

in the

Faculty of Science,

University of Cape Town.

July, 1978



Parechinus angulosus

ACKNOWLEDGEMENTS

I thank my supervisor, Professor C. von Holt, for his encouragement to keep to the task of "writing up". Also, without his efforts to provide the excellent facilities available in this department, this work would have been difficult or impossible.

I also wish to thank my husband, Associate Professor W.N. Strickland, for being available when discussion and encouragement were needed.

Dr. W.F. Brandt has also helped to further my work by helpful ideas and especially by his pioneering work to set up the sequencing methods in this laboratory.

A very profitable four month period was spent in Abteilung Wittmann (Max-Planck Institut für Molekular Genetik in West Berlin). Professor H.G. Wittmann was especially helpful in providing working space and equipment while Dr. B. Wittmann-Liebold provided many valuable discussions about sequencing techniques (e.g., solid state and micro Dansyl methods).

I wish to thank Pam Edwards for typing the manuscript and Ed Lee for photographic assistance.

General assistance by the staff of the Department of Biochemistry is gratefully acknowledged. In particular, I thank Margaret Morgan and Paul van Helden for help in collecting the many sea urchins needed for this study.

Finally, I wish to thank the Council for Scientific and Industrial Research (Chromatin Research Unit) and the University of Cape Town, whose financial assistance made this work possible.

CERTIFICATION OF SUPERVISOR

In terms of paragraph 8 of "Regulations for the Degree of Ph.D." I, as supervisor of the candidate, M. S. Strickland, certify that I approve of the incorporation in this thesis of material that has already been published or submitted for publication.

Signed by candidate

Professor C. von Holt
Head of the Department of Biochemistry

SUMMARY

Histones H2B have been purified from sperm of the sea urchin Parechinus angulosus and Psammechinus miliaris. Three H2B variants have been completely sequenced from P. angulosus and two H2B variants from P. miliaris have been partially sequenced.

The sequences of the sperm histones H2B are compared to the known sequences of other H2B histones. The sea urchin sperm histones H2B are all considerably more arginine rich than other histones H2B and contain an extended amino-terminal region containing reiterated pentapeptides.

PART 4

MATERIALS AND METHODS

PART 1

1.1 INTRODUCTION

Five major species of histones have been recognized in calf thymus chromatin and their primary structures have been determined. These five histones have been described in other species as well and in a number of cases, the primary structures have been partially or completely determined. From these investigations it appears that the sequences of histones H3 and H4 are extremely conservative over a wide spectrum of the biological world, while that of H1 is quite variable even between the H1 histones isolated from different tissues of the same organism. Histones H2A and H2B appear to be intermediate in variability. (Elgin and Weintraub, 1975).

Electrophoretic studies (4.2.1) in this laboratory of the histones associated with chromatin from mature male gonads of the southern African sea urchin Parechinus angulosus indicate the presence of proteins with similar properties to the calf thymus histones (Fig.1.1). Previous workers have interpreted similar electrophoretic findings using several species of echinoderms as proof that most somatic histones are preserved in sperm and not replaced by protamine-like proteins (Subirana and Palau, 1968, Subirana, 1970). Since electrophoretic mobility is the result of largely the size charge ratio of a protein and therefore only a general indication of the nature of a protein, a study of the sequence of purified electrophoretic components is the only sure way to draw conclusions as to the structure of a histone. For this reason a study of the primary structure of the sperm histones of P.angulosus has been in progress for some time in this laboratory.

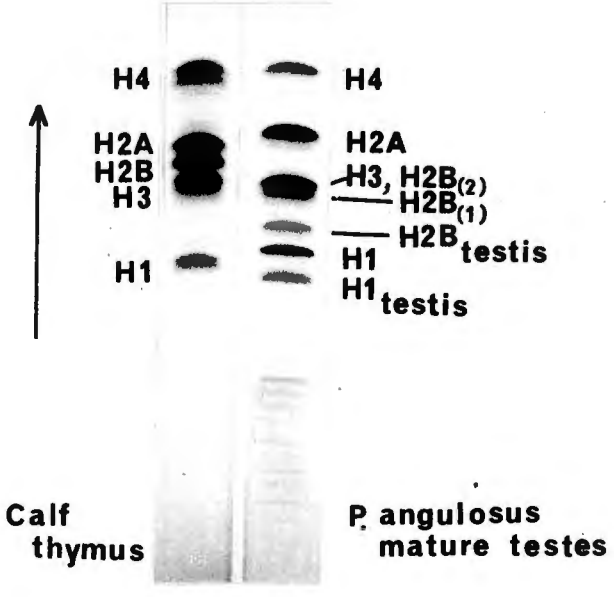


Fig. 1.1: Electrophoretic gels of whole histone extracts (0.25 N HCl) of purified nucleoprotein from calf thymus and mature testes of the sea urchin Parechinus angulosus.

One of two electrophoretic fractions in the histone H1 region (Fig.1.1) has been partially sequenced (Strickland et.al., 1976). This particular fraction is the only component in the histone H1 region when sperm cells are the source of nucleoprotein and has been named sperm histone H1. The other slower moving band which is present only when whole testes are extracted has been called testis histone H1, but has not been further studied. The amino acid composition of sperm histone H1 is similar to a sperm-specific histone called γ which is rich in lysine, arginine and alanine isolated by Paoletti and Huang (1969)from another sea urchin, Arbacia punctulata.

Histone H3 from sperm cells of P.angulosus has an identical electrophoretic mobility and a similar amino acid composition to calf thymus histone H3. The sea urchin histone H3 has one cysteine while calf histone H3 has two. The first 48 amino-terminal residues are identical in histone H3 of sea urchin and in histones H3 isolated from calf thymus, chicken erythrocytes, shark erythrocytes and mollusc testes. Histone H3 from cycad pollen had a single change in the 48 N-terminal residues examined; at position 39 phenylalanine replaces the tyrosine found in histone H3 from the five other sources just mentioned. (Brandt et al., 1974).

Histone H4 from sperm of P.angulosus has an amino acid composition differing from calf in having one less threonine and one more cysteine. Partial sequence of the 52 carboxy-terminal amino acids of histone H4 from Parechinus revealed cysteine at position 73 where threonine is found in calf H4. (M.Strickland et al., (1974). Similar results have been reported for H4 from gonads of the sea urchin Psammechinus miliaris (Wouters-Tyrou et al.,(1976).

The structure of H2A from sea urchin sperm is still under investigation. The amino-terminus of this protein is blocked and the amino acid composition has similarities to that of H2A from calf thymus tissue. A partial sequence of an N-bromosuccinimide peptide shows strong homologies with the calf thymus H2A. (Brandt et al., manuscript in preparation).

Histone H2B from calf thymus is classified as a lysine rich histone and migrates on acrylamide gels between H3 and H2A. A protein of a similar composition or electrophoretic mobility cannot be found in an acid extract from chromatin of sperm cells of P. angulosus. Instead, three different arginine rich proteins with electrophoretic mobilities similar to calf histone H3 have been isolated and classified as histones H2B. (W.N. Strickland et al., 1974; M. Strickland et al., 1977; W.N. Strickland et al., 1977; and M.Strickland et al., 1978a). When acid extracts are made from chromatin from whole testes, a slower moving fraction between the sperm H2B histones and sperm H1 (Fig. 1.1) is present which appears to be a mixture of proteins similar to the arginine rich H2B proteins present in sperm (amino acid analysis - Table 1.1 - and partial sequence). This fraction has been called testis H2B but has not been studied further.

Two arginine rich histones H2B, also with electrophoretic mobilities near calf histone H3, have been isolated from sperm cells of a European species of sea urchin, Psammechinus miliaris and partially sequenced. (M. Strickland et al., 1978b). The investigation of the structures of these five sperm specific histones H2B from two species of sea urchin is the subject of this dissertation.

	Heterogeneous sperm H2B	Heterogeneous testis H2B	H2B calf*
Aspartic Acid	3.6	3.9	4.8
Threonine	7.4	6.6	6.4
Serine	9.5	8.4	11.2
Glutamic acid	6.6	7.8	8.0
Proline	5.8	5.2	4.8
Glycine	9.1	9.2	5.6
Alanine	7.2	9.5	10.4
Valine	6.8	7.3	7.2
Methionine	1.5	1.5	1.8
Isoleucine	3.3	3.6	4.8
Leucine	4.5	4.9	4.8
Tyrosine	3.4	2.9	4.0
Phenylalanine	1.7	1.7	1.6
Lysine	11.0	11.1	16.0
Histidine	1.5	1.5	2.4
Arginine	17.1	14.8	6.4

Table 1.1: Amino acid compositions (mole per cent) of heterogeneous sperm and testis H2B.

These two fractions were separated after Biogel P60 chromatography on which the testis H2B elutes just previous to sperm H2B. Heterogeneous sperm H2B shows two electrophoretic bands near H3 which have been separated and sequenced ($H2B_{(1)}$ and $H2B_{(2)}$ *Parechinus* Fig 1.4, Part 2.2 and 2.3).

The testis fraction showed one electrophoretic band between H1 and H3 but contained two amino terminal sequences (10 steps) consistent with a mixture of $H2B_{(1)}$ and $H2B_{(2)}$ *Parechinus*.

*Data from Iwai et al., 1972.

1.2 PREPARATION AND PURIFICATION OF H2B HISTONES FROM SPERM CELLS OF PARECHINUS ANGULOSUS AND PSAMMECHINUS MILIARIS

1.2.1 Preparation of crude H2B by selective extraction

Sperm of Parechinus angulosus collected at locations on the eastern side of the Cape Peninsula were used in all initial experiments on purification of the H2B proteins.

Differences encountered when collections were made on the western side of the Cape Peninsula and when sperm of the European species of sea urchin Psammechinus miliaris were used will be mentioned after the discussion of purification.

The different binding properties of histones to DNA and their different solubilities in aqueous and organic solvents are exploited in the selective extraction method of Johns (1964). Part of this technique in conjunction with exclusion (4.1.2.1) and ion exchange (4.1.2.2) chromatography has been used to purify sea urchin histones.

After initially washing sperm nucleoprotein with 0,15 M NaCl, most of the H1 was removed with dilute perchloric acid (4.1.1). H3, H2A and H4 were next extracted with ethanolic - HCl (4.1.1) and finally the H2B histones with small amounts of the other histones and some non-histone proteins were extracted with 0.25N HCl (4.1.1). An acid extract (0.25N HCl) of the washed nucleoprotein contains all the histones. Figure 1.2 shows the electrophoretic pattern of a whole histone extract compared to histones obtained by selective extraction. The H2B region

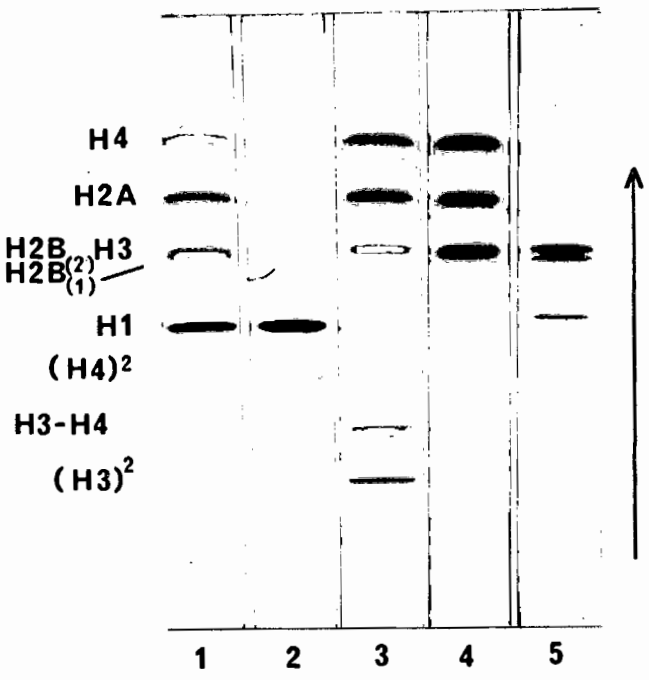


Fig. 1.2: Electrophoretic gels of samples from various stages in the selective extraction (4.1.1) of sperm nucleoprotein of Parechinus angulosus compared to a total extract from nucleoprotein. Collections were from the eastern side of the Cape Peninsula.

(1) whole histone extract with 0.25 N HCl;

(2) perchloric acid extraction;

(3) ethanolic-HCl extraction;

(4) ethanolic - HCl extraction + DTE as a reducing agent;

(5) HCl extraction.

on gel 5 is seen to contain two bands. Typical yields after selective extraction of 100 g wet packed sperm were: 0.7 g (histone H1) in the perchloric acid extract, 1.7 g (histone H4, H3, H2A) in the 80 % ethanol-0.25 N HCl extract and 1.0 g (histone H2B , H1 and non-histone proteins) in the 0.25 N HCl extract; about 0.5 g of H2B mixture (Fig.1.4 gel 2) was obtained from the 0.25 N HCl extract after Biogel P60 fractionation. The yield of histone H2B was 17.7 % of the total histones extracted.

1.2.2 Column Chromatography of Crude H2B

1.2.2.1 Gel exclusion chromatography on Biogel P60

Chromatography on Biogel P.60 (4.1.2.1) separated a major protein peak with two electrophoretic components (Fig. 1.3A, Fig.1.4, gel 2). Because the only amino end group found by dansylation (4.2.2.1) of this fraction was proline, as is the case for calf histone H2B, this P60 fraction was presumed to contain histone H2B.

Histone H3 co-electrophoresis with one of the two histone H2B bands but would have been detected by its alanine amino-end group. Amino acid analysis of the histone H2B fraction indicated some similarities to histone $H2B_{calf}$ but also a considerable increase in arginine content in the sea urchin H2B fraction (Table 1.1). To rule out the possibility that the two electrophoretic components represented post-synthetically modified forms of a single protein, the histone H2B fraction from Biogel P 60 was subjected in a pilot experiment to sequence analysis (4.4.3). It was apparent from the results that two distinct proteins both with amino-terminal proline were present. Therefore, some further purification was obviously necessary.

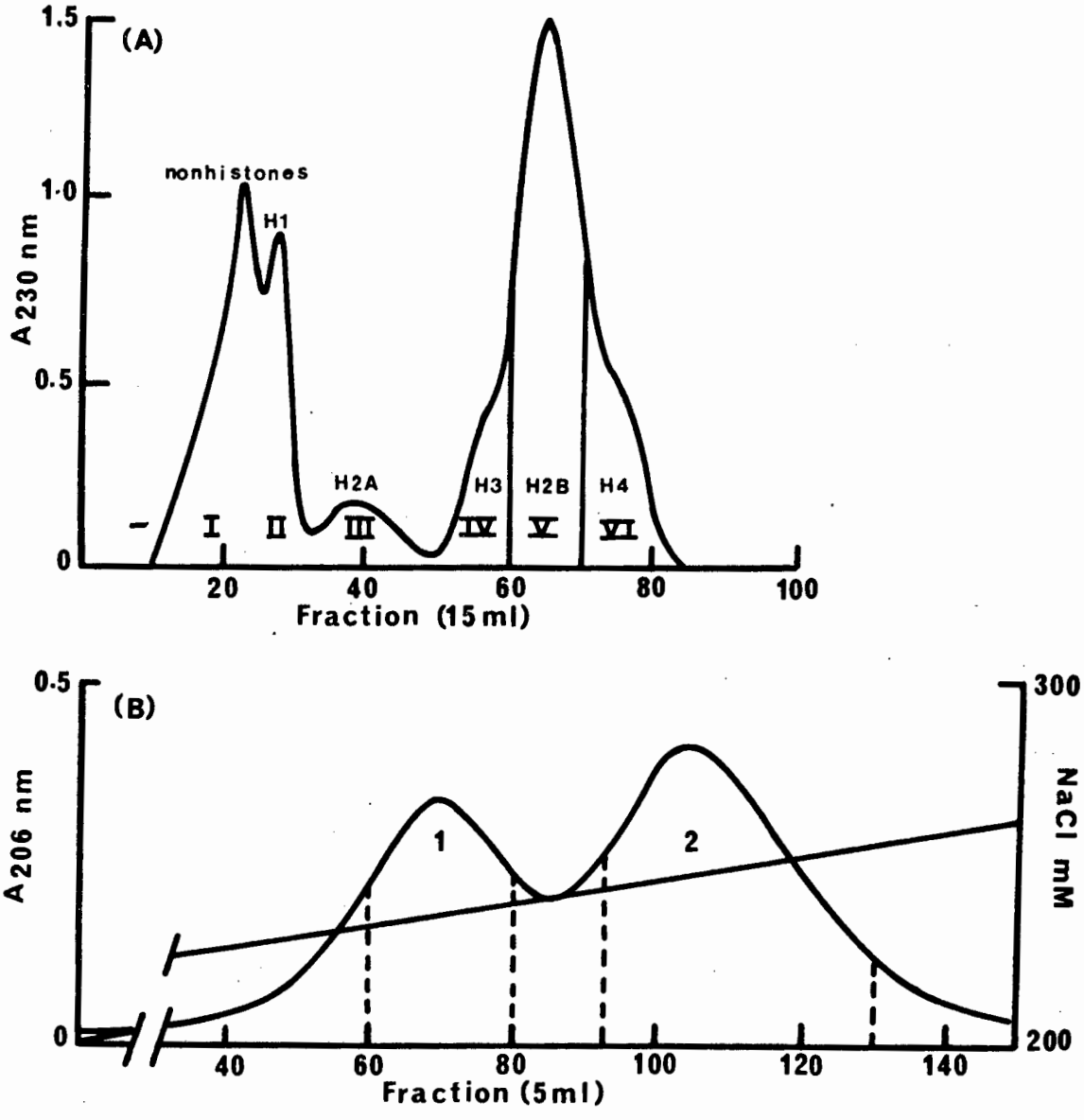


Fig. 1.3: Elution patterns of (A) = selectively extracted crude H2B from sperm of P. angulosus chromatographed on a 50 x 900 mm Biogel P60 column with 10 mM HCl as the eluant. (B)= 130 mg fraction V from the P60 column chromatographed on a 26 x 300 mm CMC column. Equilibration buffer was 50 mM sodium acetate/HCl, pH 4.5, 6M urea; linear gradient: 200 - 300 mM NaCl in equilibration buffer, total volume 700 ml. Flow rate: 40 ml/hour. Yields: 1 = 25 mg 2 = 40 mg.

Source of sample was sperm of Parechinus angulosus collected from the eastern side of the Cape Peninsula.

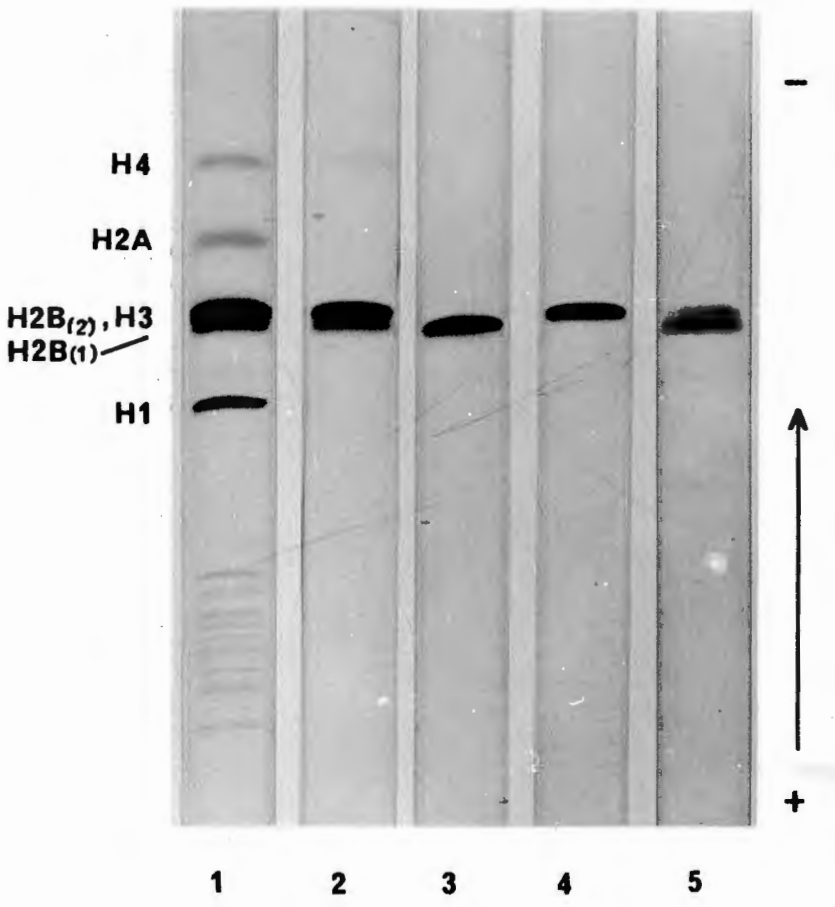


Fig. 1.4: Polyacrylamide electrophoresis at various stages in the purification of H2B histones from sperm of P. angulosus collected on the eastern side of the Cape Peninsula.

(1) selectively extracted crude H2B (0.25 N HCl);

(2) H2B after Biogel P60 chromatography (fraction V, Fig.1.3A) of crude H2B;

(3) $H2B_{(1)}$ after CMC chromatography (fraction 1, Fig. 1.3B) of 2 above;

(4) $H2B_{(2)}$ after CMC chromatography (fraction 2, Fig. 1.3B) of 2 above;

(5) mixture of 3 and 4 = 4:1.

1.2.2.2 Ion exchange chromatography on carboxymethyl cellulose

Ion exchange chromatography on a weak cation exchange resin such as CMC seemed a method suitable for separating the two H2B proteins. We had previously shown that a mixture of the CNBr (4.3.1.1) amino terminal peptides generated from a mixture of these proteins separated on CMC using a NaCl gradient and an acetate buffer at pH 4.5 (4.3.3.2) Strickland, W.N.et al.,1974). The intact proteins, however, could not be eluted from CMC even by 1 M NaCl. Only when 6 M urea was present in the acetate buffer could two protein fractions with distinct electrophoretic mobilities be eluted from CMC using a NaCl gradient of 200-300 mM (4.1.2.2, Fig. 1.3B, Fig. 1.4, gels 3-5). The two proteins were called $H2B_{(1)}$ Parechinus and $H2B_{(2)}$ Parechinus according to their elution order from the CMC column.

Electrophoretic studies on the histone H2B fraction from sperm of the same species (P.angulosus) collected on the western side of the Cape Peninsula, indicated differences in the electrophoretic pattern compared to those fractions derived from organisms collected on the eastern side of the Cape Peninsula. (Compare Fig. 1.5A, gel 1 and Fig.1.4, gel 2). In the former material the slower moving band of the doublet appeared more intense. The intensity of this band varied with the location site. In material collected at 12 meters depth (Fig. 1.5A, gel 1) it was the major band while in collections made at the intertidal zone the two bands stained with about equal intensity. (Fig.1.5B, gel 1). In contrast collections from the eastern side of the peninsula have consistently yielded preparations in which the faster moving band is equal or darker than the slower one. All this could simply reflect a difference in the amount of $H2B_{(1)}$ and $H2B_{(2)}$ or

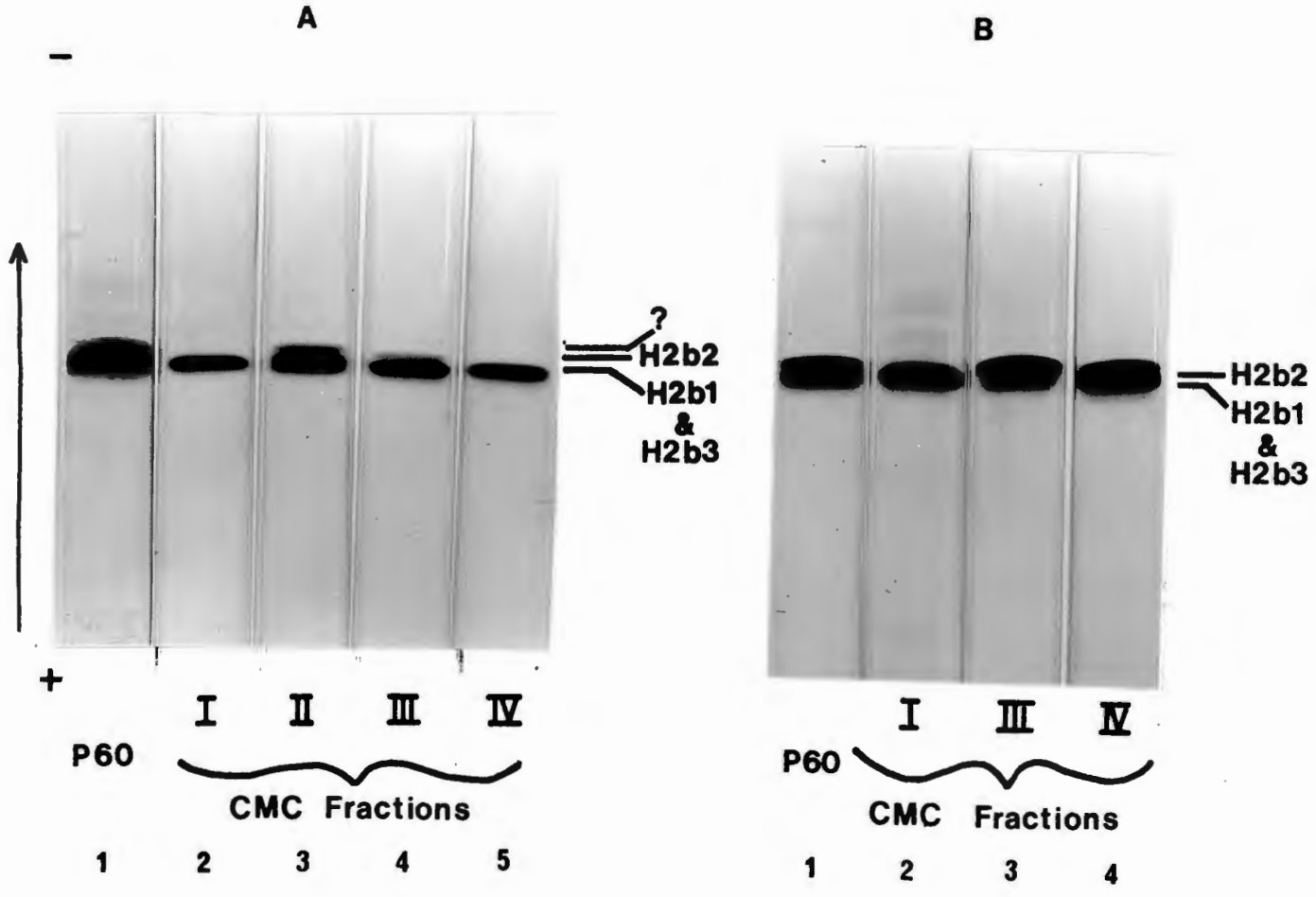


Fig. 1.5: Polyacrylamide electrophoresis at various stages of purification of H2B histones from sperm of P. angulosus collected on the western side of the Cape Peninsula.

P60: Crude H2B fraction recovered from exclusion chromatography on Biogel P60.

CMC - Fractions: recovered after ion exchange chromatography of P60 fraction.

A: Histones from animals collected at 12 meters depth (Fig. 1.6).

B: Histones from animals collected at the intertidal zone. (Fig. 1.7).

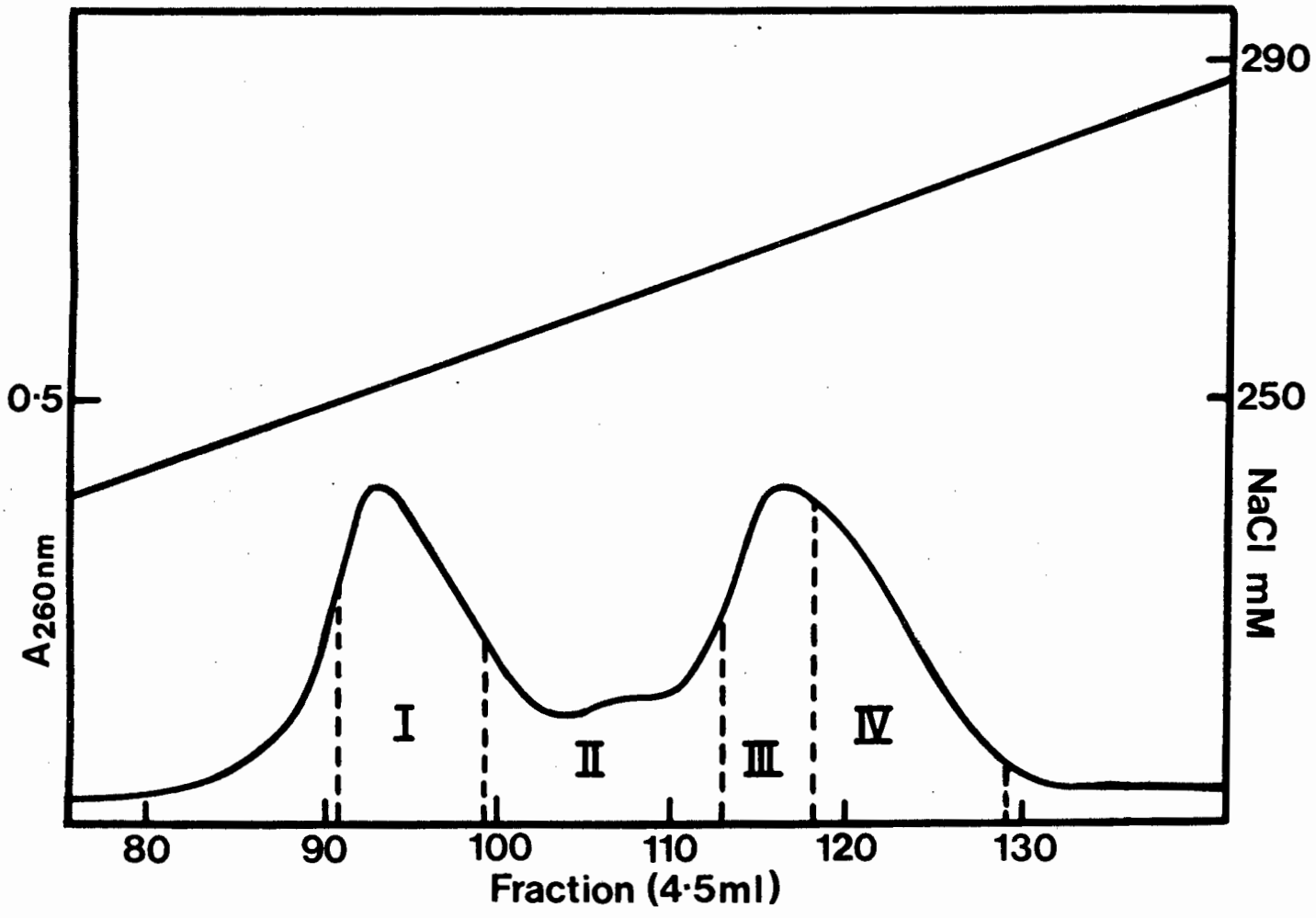


Fig. 1.6: Elution pattern from CMC of Biogel P60 H2B histone fractions isolated from sperm of P. angulosus collected from 12 metres depth. 127 mg applied to column, yield : I 30 mg, II 17 mg, III 30 mg, IV 31 mg. Column 26 x 400 mm. Cuvette 0.5 mm. Linear gradient of 200-300 mM NaCl in 50 mM sodium acetate/HCl pH 4.5 and 6 M urea. Total volume 700 ml, flow rate:35 ml/hour. See Fig. 1.5Afor electrophoretic gels.

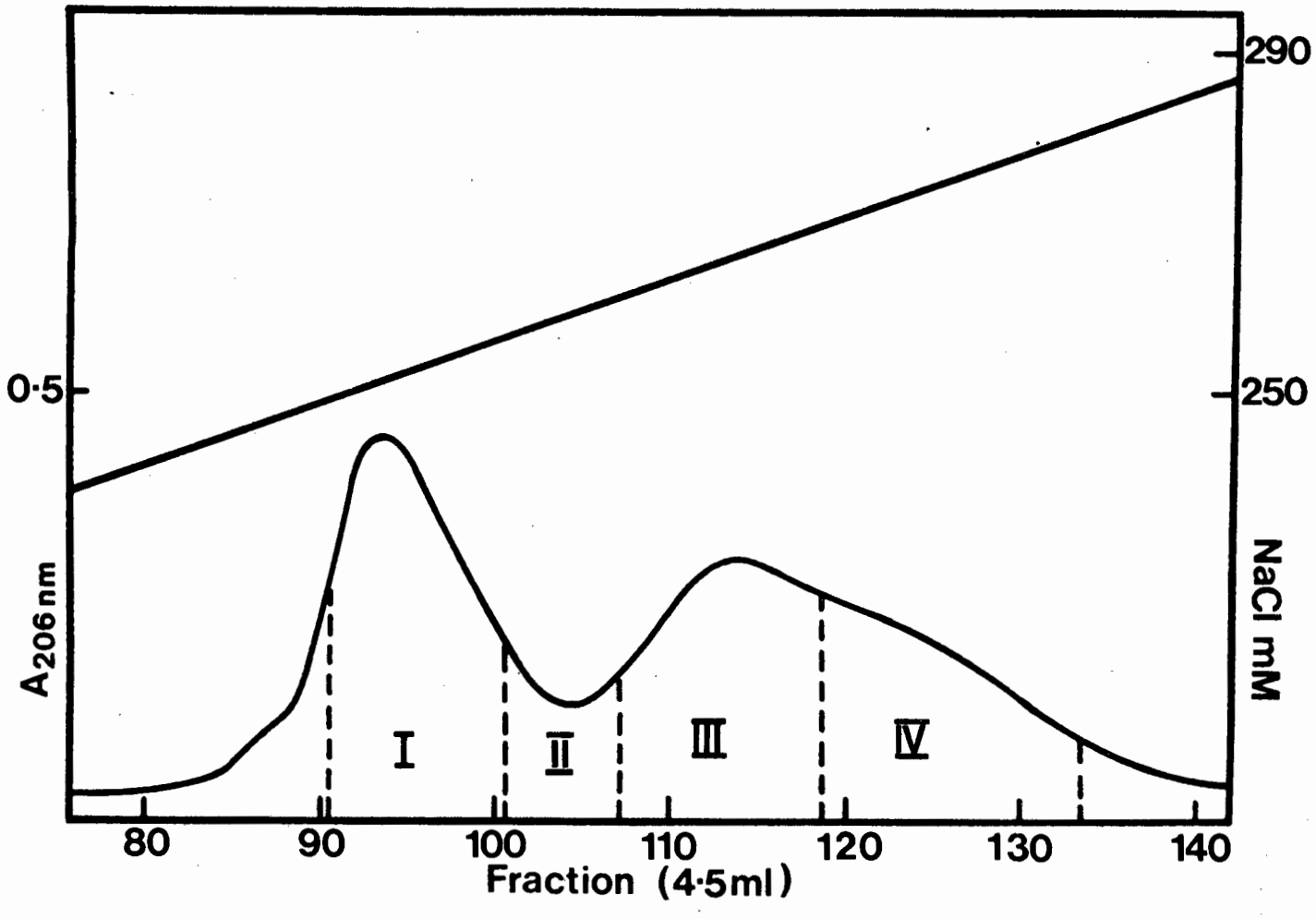


Fig. 1.7: Elution patterns from CMC of Biogel P60 H2B histone fractions isolated from sperm of P. angulosus collected from the intertidal zone of the western side of the Cape Peninsula.
Conditions for chromatography are the same as for Fig. 1.6.
See Fig. 1.5B for electrophoretic gels.

a difference in the amount of contamination by H3. However, after CMC chromatography it became apparent that a new protein fraction was present which eluted just after $H2B_{(2)}$ (Fig. 1.6 and Fig. 1.7) but co-electrophoresed with $H2B_{(1)}$ (Fig. 1.5A and B). This protein, called $H2B_{(3)}$ Parechinus was purified from sea urchins collected from 12 meters depth where it almost completely replaced $H2B_{(2)}$ (Fig. 1.5A, gel 5, Fig. 1.6 fraction IV).

Sperm of the European species of sea urchin Psammechinus miliaris was provided by Dr. M.L. Birnstiel and was extracted as described for P. angulosus. Electrophoretic studies of the selectively extracted crude H2B (0.25 N HCl) extract showed a single band in the H2B region which migrated slightly faster than the H2B doublet of P.angulosus (Fig.1.8) Biogel P60 fractionation yielded a protein fraction with a single electrophoretic component (Fig. 1.8 gel 4). This fraction had proline as the sole amino group detected by dansylation. CMC chromatography yielded two peaks, each of which had an identical mobility to the P60 starting material (Fig. 1.9).

Wouters - Tyrou (1977) reports purification of a gonad H2B from P. miliaris with an electrophoretic mobility between H3 and H1 and with N-terminal proline and C-terminal arginine. This fraction may be similar to the testis H2B isolated from P. angulosus (Table 1.1, Fig. 1.1).

1.3 CHARACTERIZATION OF H2B HISTONES

1.3.1 End group analysis

All five sea urchin sperm H2B histones have amino terminal proline (Dansylation 4.2.2.1) as does H2B from calf thymus.

Carboxypeptidase A and B digestion (4.2.2.2) yielded arginine as the carboxyl terminus of all sperm H2B proteins (Fig. 1.10, 1.11, 1.12) while the calf H2B has carboxy-terminal lysine (Iwai et al., 1972).

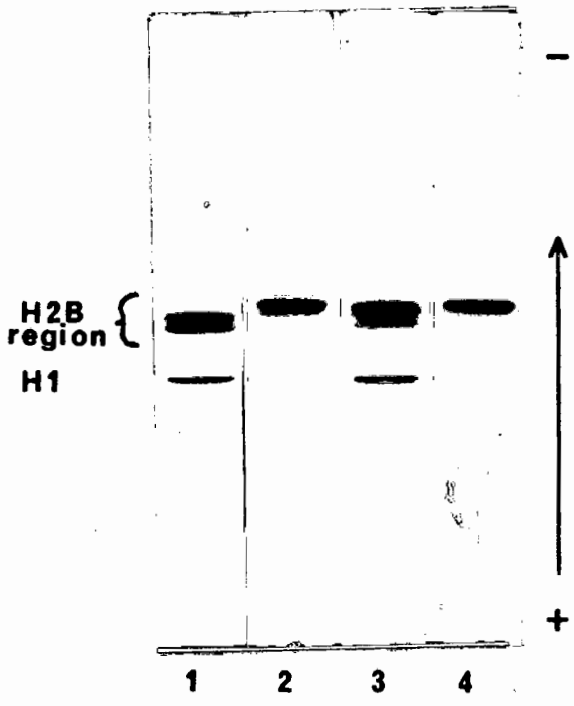


Fig. 1.8: Polyacrylamide elctrophoresis comparing selectively extracted crude H2B from sperm of Parechinus angulosus and Psammechinus miliaris.

1 = Crude H2B P. angulosus; 2 = Crude H2B P. miliaris;
3 = Mixture of 1 and 2; 4 = H2B fraction of P. miliaris after Biogel P60 chromatography.

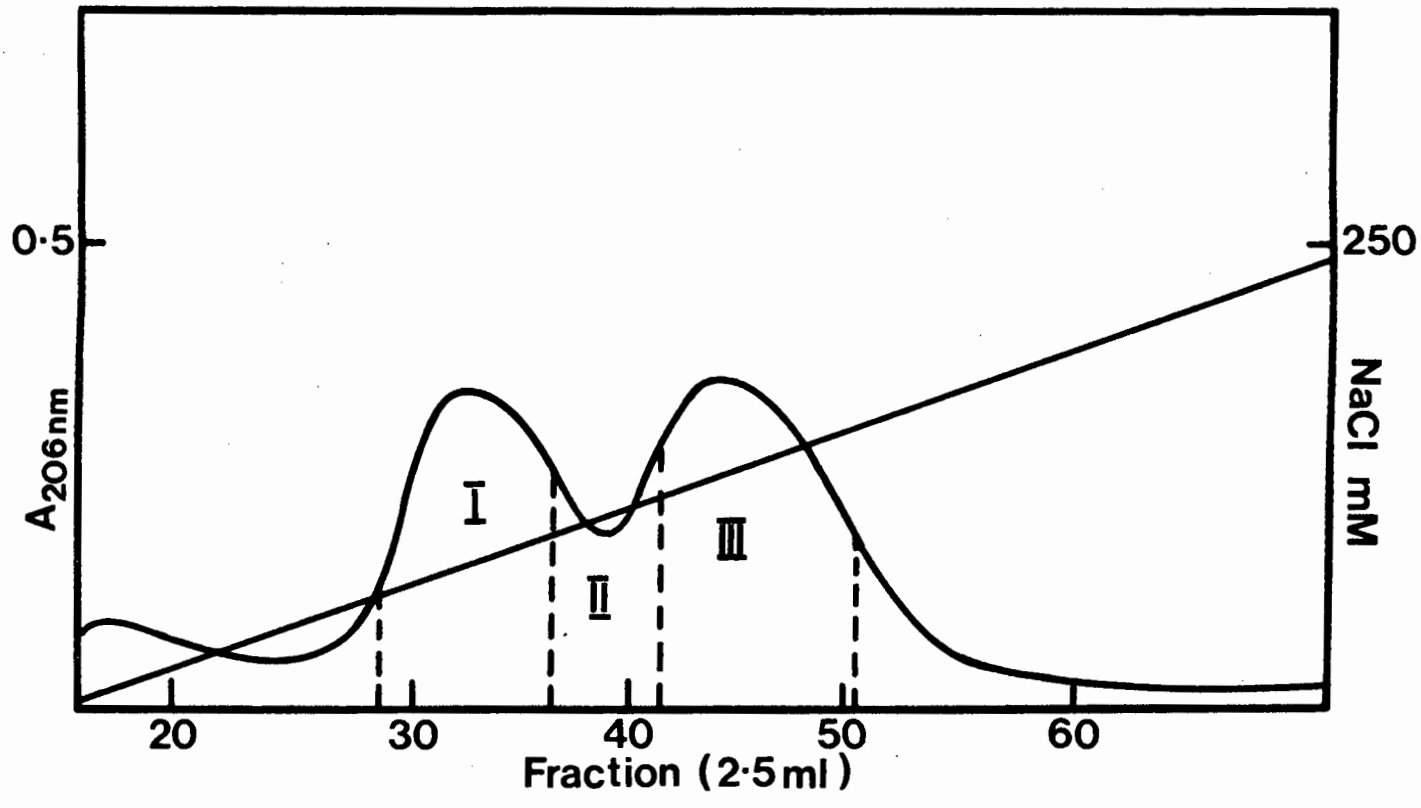


Fig. 1.9: Elution pattern from CMC of Biogel P60 H2B histone fraction isolated from sperm of Psammechinus miliaris.

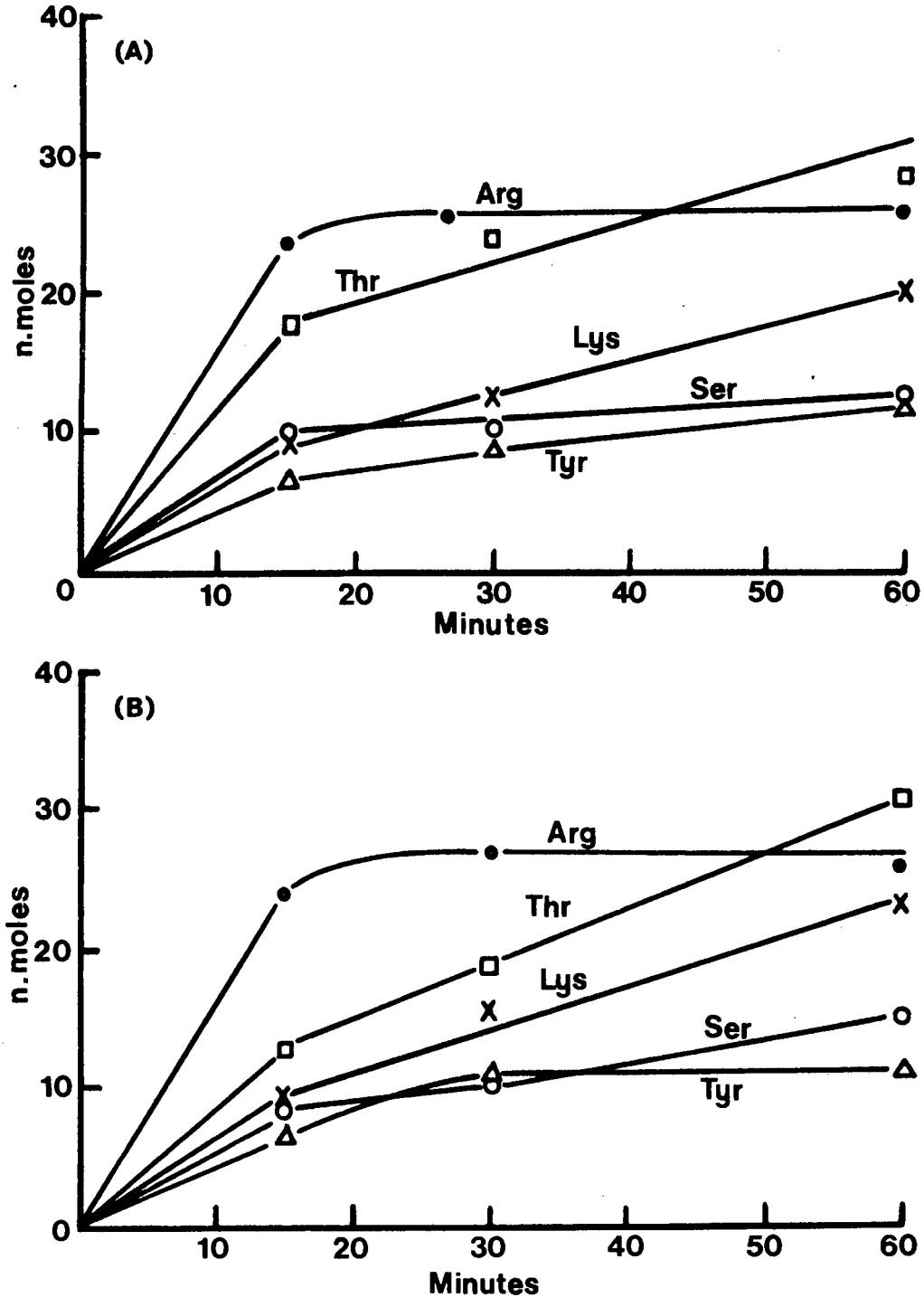


Fig. 1.10: Yields of amino acids after incubation for various times of $H2B_{(1)}$ (A) and $H2B_{(2)}$ (B) from Parechinus angulosus with a 1 : 100 molar ratio (enzyme: substrate) carboxypeptidase A and a 1:200 molar ratio of carboxypeptidase B. Approximately 35 n moles of H2B protein was in each sample tested (4.2.2.2).

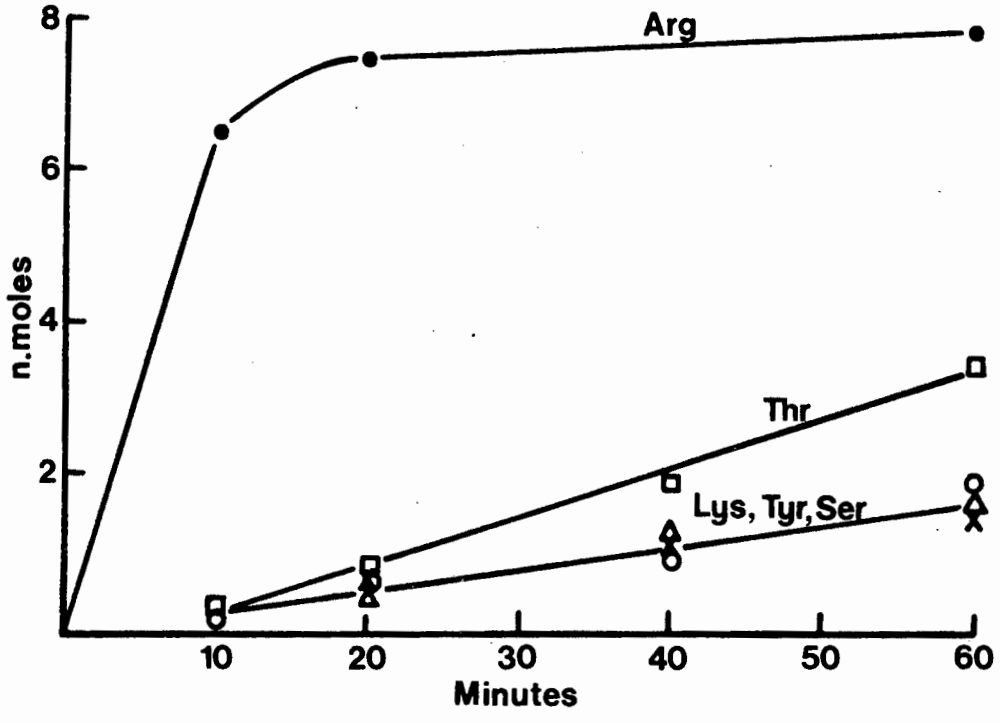


Fig. 1.11: Yields of amino acids after incubation of $H2B_{(3)}$ from Parechinus angulosus for various times with a 1:500 molar ratio (enzyme: substrate) of Carboxypeptides A and B (4.2.2.2). Each sample contained approximately 10 nano moles of $H2B_{(3)}$.

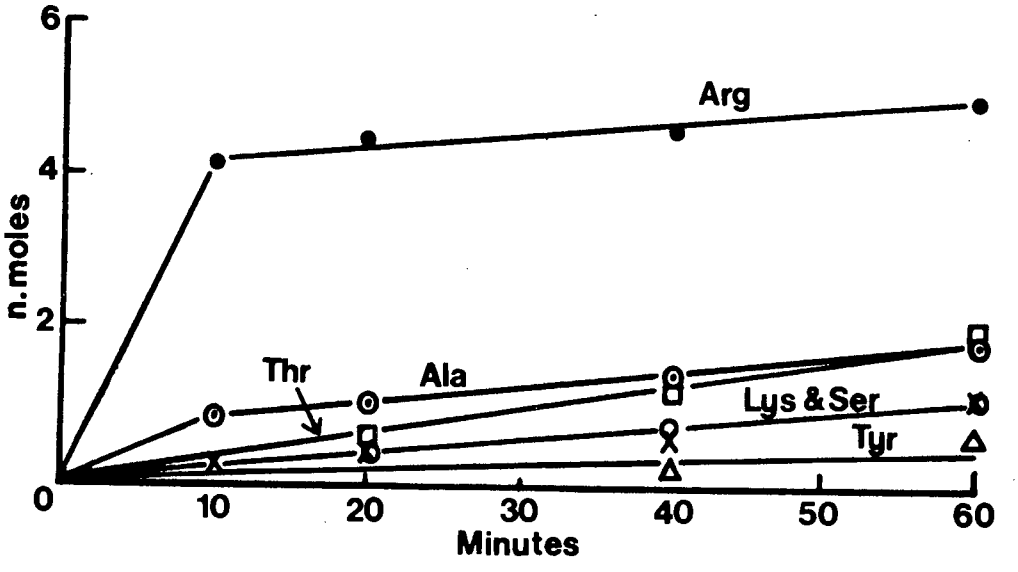


Fig. 1.12: Yields of amino acids after incubation of $H2B_{(2)}$ from Psammechinus miliaris with a 1/500 molar equivalent of carboxypeptidases A and B (4.2.2.2). About 7 nano moles of protein $H2B_{(2)}$ was present in each sample tested. The yields in nano moles of amino acids at 40 minutes incubation for $H2B_{(1)}$ under the same conditions as above were: threonine 2.7, serine 1.5, alanine 0.6, tyrosine 0.8, lysine 1.5, and arginine 4.0.

1.3.2 Amino acid composition

The amino acid composition after acid hydrolysis (4.2.3) of the five sperm H2B histones is compared to that of the calf thymus H2B in Table 1.2. It can be seen that there are similarities in general composition but that the arginine content is higher in all the sperm proteins.

	$H2B_{(1)}$ P. miliaris	$H2B_{(1)}$ P.angulosus	$H2B_{(2)}$ P. miliaris	$H2B_{(2)}$ P.angulosus	$H2B_{(3)}$ P.angulosus	H2B * calf
Aspartic acid	6.0	4.7 (5)	6.0	4.8 (5)	4.9 (5)	(6)
Threonine	9.9	11.2 (12)	7.0	7.8 (9)	7.4 (9)	(8)
Serine	12.9	13.3 (15)	15.7	15.7 (18)	17.2 (20)	(14)
Glutamic acid	10.2	11.2 (11)	8.2	8.3 (8)	8.2 (8)	(10)
Proline	5.9	6.1 (7)	8.3	7.7 (8)	9.2 (9)	(6)
Glycine	12.2	14.6 (14)	15.1	13.1 (13)	13.0 (13)	(7)
Alanine	10.2	10.1 (10)	10.4	11.8 (12)	12.0 (12)	(13)
Valine	12.2	11.6 (12)	9.8	10.6 (11)	10.4 (11)	(9)
Methionine	1.8	1.7 (2)	1.8	2.0 (2)	1.8 (2)	(2)
Isoleucine	4.0	3.5 (4)	5.9	4.8 (5)	4.8 (5)	(6)
Leucine	6.1	5.8 (6)	6.2	6.1 (6)	6.1 (6)	(6)
Tyrosine	4.9	4.4 (5)	3.6	4.1 (4)	3.9 (4)	(5)
Phenylalanine	2.1	1.6 (2)	2.0	2.2 (2)	2.0 (2)	(2)
Lysine	14.4	15.8 (16)	15.8	14.1 (14)	15.2 (15)	(20)
Histidine	2.1	1.7 (2)	1.9	1.8 (2)	1.9 (2)	(3)
Arginine	20.2	22.8 (21)	23.8	24.8 (24)	25.2 (25)	(8)

Table 1.2: Amino acid compositions in molar ratios of three H2B histones from P. angulosus and two H2B histones from P. miliaris compared to that of the H2B from calf thymus tissue.

Figures in parentheses are residues determined from sequence analysis. No corrections have been made for losses and incomplete cleavage during acid hydrolysis.

* Data from Iwai et al., 1972.

PART 2

2.1 INTRODUCTION

Modern structural protein chemistry is based on the sequential N-terminal degradation as developed by Edman (1950). This method is suitable for the degradation of oligo- and polypeptides. Depending on the nature of the material to be sequenced, a choice may be made from several variations of the basic methodology.

The number of degradation steps which can be achieved has been considerably extended by the introduction of an automatic protein sequenator (Edman and Begg, 1967) in which peptides are degraded in solution in the "spinning cup" method. Alternative methods based on the covalent linkage of the peptide to a solid phase (Laursen, 1971) are particularly useful when the peptide cannot be retained as a stable film in the spinning cup sequenator. The micro DNS-Edman sequencing method (Chen, 1976. Narita, 1975) is very suitable when small amounts (< 50 n moles) of a small peptide are to be sequenced. This method is extremely sensitive but does not lend itself to the quantitative determination of the amino acid residues exposed after degradation.

In any of these methods the number of repetitive degradations which can be achieved is limited in practice due to incompleteness of the reaction and non-specific cleavage of the protein in the chemicals used. Therefore, fragmentation of the protein is necessary and the generation and purification of peptides usually forms the major task in sequence determination.

The five proteins which have been completely or partially sequenced in this study have similar structural features and certain fragmentation techniques have been used on more than one protein. Consequently, each cleavage technique and the purification of the peptides generated, will be described

for one protein and not repeated if similar conditions and results were obtained with the other proteins. The amino acid sequences of peptides generated have been determined as the case may be either by the spinning cup method, a solid state methodology or by the manual DANSYL-Edman method metnioned previously.

2.1.1 Strategy of sequence determination

The first objective is to fragment the protein into a minimum of large overlapping or adjacent peptides by chemical and enzymatic cleavages. The choice of the methods to be employed will be guided by the amino acid composition. In general, the fewer the cleavage points the simpler the purification of the peptides. Reasonable predictions as to the type of peptides generated could be made for the carboxy-terminal regions of the proteins once their homology to the H2B from calf thymus for which the complete sequence is known had been established. (Iwai, 1972). The amino ends, however, of the sea urchin sperm H2B histones were completely different from the amino terminus of the calf protein and predictions could only be made after pilot amino terminal sequencing and study of the amino acid composition of the amino terminal CN-I peptides. In most cases no attempt was made to purify and fully characterize all the peptides generated by a particular cleavage method. Instead certain fractions from chromatography were selected for further purification after examination of characteristics such as apparent size, ultraviolet absorbance at 280 (tyrosine or phenylalamine present), DNS amino groups present, and amino acid composition.

The H2B proteins in this study contain two methionine residues. Cyanogen bromide is known to specifically cleave at this residue (Gross, 1967; Witkop, 1968). CNBr cleavage did yield two large peptides from all the H2B histones studied and these peptides were sequenced and also used as subjects for further fragmentation.

Trypsin specifically cleaves at the carboxyl side of arginine and lysine residues (Smyth, 1967). Reaction of the ε-amino groups of lysine with maleic anhydride (Butler and Hartley, 1972) restricts the cleavage points to arginine residues. In view of the high content of basic amino acids in the N-terminal half, this cleavage proved useful only in the C-terminal half of the molecule.

A protease from Staphylococcus aureus (Houmard and Drapeau, 1972; Drapeau, 1976) with a cleavage specificity for glutamic acid has proved useful in this study to produce peptides following isolated glutamic acid residues in the center of the molecule and the carboxy terminal region.

Chymotrypsin is not as specific an enzyme as trypsin and causes cleavage at the carboxyl side of tyrosine, phenylalanine and tryptophan (Smyth, 1967). It is reported that cleavage at methionine is slow but in this study complete cleavage at methionine was observed after short incubation times. Chymotrypsin was used when cleavage at tyrosines was desired.

Thermolysin has a rather non-specific cleavage capacity (Matsubara, 1970). Hydrophobic residues such as valine, isoleucine, leucine, phenylalanine, alanine, methionine and tyrosine are cleaved at their amino terminal bond. This enzyme was usually used as the final cleavage agent at isolated hydrophobic residues.

Cleavage at aspartic acid residues in dilute acid (Schultz, 1967) was used in cases where the desired peptide could not be obtained by any of the methods mentioned previously. This method does produce peptides resulting from cleavage at aspartic acid residues but considerable non-specific cleavage occurs resulting in a low yield of the desired peptide.

2.1.2 Purification of peptides and nomenclature

The first two letters in the code for a peptide indicate the type of cleavage by which the peptide had been generated (CN = cyanogen bromide, TH = thermolysin, SP = Staphylococcus aureus protease, MT = tryptic digestion of maleylated protein, CH = chymotrypsin, AS = dilute acid hydrolysis at aspartic acid). The numbers which follow indicate the elution order on column chromatography. A Roman numeral refers to separation by gel filtration on Sephadex (4.3.3.1) an Arabic number indicates chromatography on CMC (4.3.3.2). For example the peptide CN-I, TH-II-10 from $H2B_{(3)}$ resulted from cyanogen bromide cleavage of the intact protein and subsequent recovery of the first peptide which eluted from a Sephadex G-100 column (Fig.2.1). This CN peptide was digested with thermolysin and the peptide in question was recovered after Sephadex G-50 chromatography in an impure fraction which eluted as the second fraction (Fig.2.14A). This fraction was further subjected to CMC chromatography and the pure peptide eluted as the tenth fraction (Fig.2.14B). (See also Table 2.5).

Peptides from the CMC columns were desalted depending on size, on Sephadex G-10, G-25 or G-50 columns with 10 mM HCl as the eluant. Steps on desalting columns were not included in the coding of the peptide unless some fractionation was achieved. Typical desalting runs are illustrated in Fig.2.6. Homogeneity of a peptide at each stage of purification was judged in the first instance by determination via dansylation of the N-terminal amino groups present (4.2.2.1). If only one amino acid was present the preparation was presumed to be pure and subjected to amino acid analysis and finally to sequence analysis.

2.2 SEQUENCE OF $H2B_{(2)}$ PARECHINUS

2.2.1 Generation and purification of peptides

2.2.1.1 Cyanogen Bromide cleavage

The whole protein was cleaved with CNBr (4.3.1.1) and the peptides fractionated on Sephadex G-100 (4.3.3.1). Figure 2.1 shows a typical separation. See also Table 2.1 for amino acid composition of peptides.

2.2.1.2 Trypsin digestion

Trypsin was used to digest maleylated CN-II which represented the carboxyl 63 amino acids of the protein (4.3.2.1). Peptide CN-II, MT-I was recovered from a Sephadex G-50 column (Fig. 2.2, Table 2.1).

In experiments where the whole protein was digested, there was evidence that some cleavage had occurred at methionine; the expected 15 amino acid peptide spanning the two methionines was never recovered and methionine was found (by amino acid analysis) in the very small (2-3 amino acid) peptide fraction. This cleavage at methionine could be a result of contamination of the trypsin with a chymotryptic type of activity.

Another reason for non-recovery of this peptide may be that being very hydrophobic it is consequently insoluble in the 10 mM HCl used when the medium-sized peptide fraction from the G-50 column (eluant-100 mM NH_4HCO_3 pH 8.5) was re-run on G-25 or G-10. Use of acetic or formic acid eluants might make recovery of this peptide possible. Detection methods for peptides which are not dependant on a change in absorption in the ultraviolet range are necessary when acetic or formic acid are used as eluants since these compounds have a very high absorbance in this region.

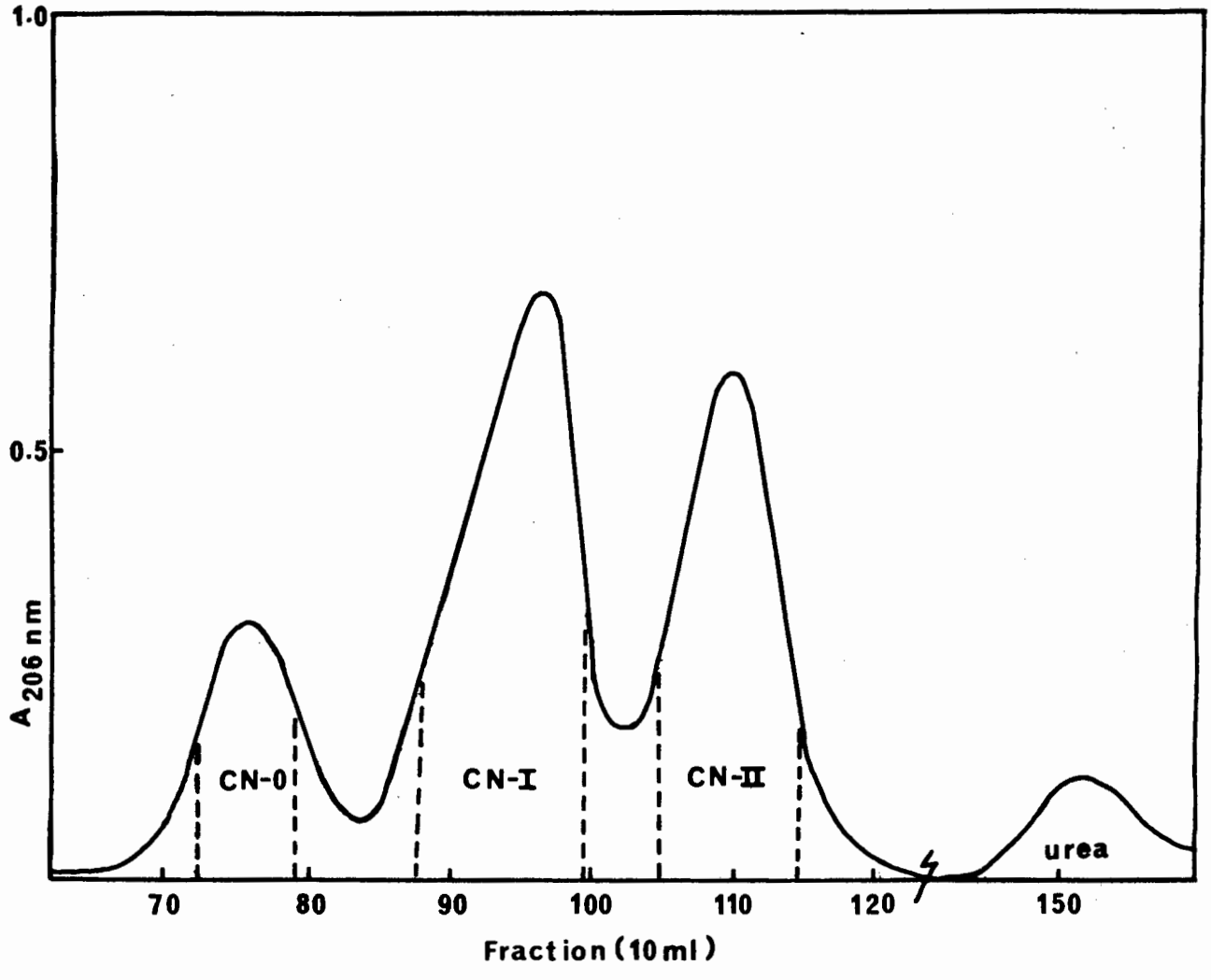


Fig. 2.1: Fractionation of a CNBr digest of H2B(2) Parechinus
Column 50 x 1000 mm Sephadex G-100, eluant: 10 mM HCl
cuvette: 1 mm, sample: 68 mg applied in 2 ml 8 M urea in
eluant, yield: CN-0 6 mg, CN-I 26.4 mg, CN-II 18.5 mg.

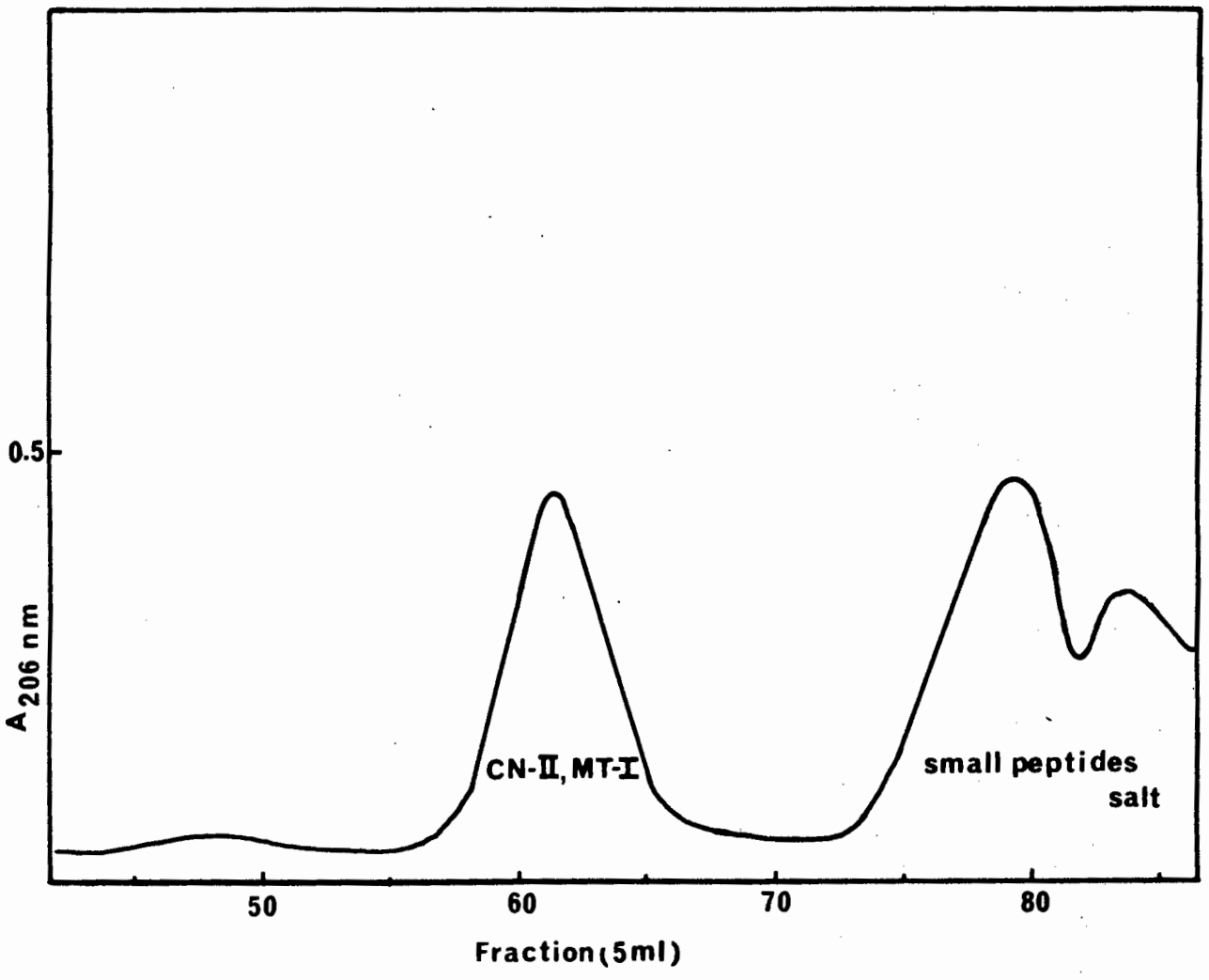


Fig. 2.2: Fractionation of a tryptic digestion of CN-II of H2B(2) Parechinus Column 26 x 1000 mm Sephadex G-50 medium, eluant: 100 mM NH_4HCO_3 pH 8.5, cuvette: 1 mm, sample 30 mg applied in 1 ml eluant, yield: CN-II, MT-I 7.8 mg.

2.2.1.3 Cleavage with Staphylococcus aureus protease

CN-I was incubated with S.aureus protease at pH 4.0 in 50 mM ammonium acetate buffer (4.3.2.2). Cleavage was nearly complete and two pure peptides (CN-I, SP-I and CN-I, SP-II) were obtained after fractionation on G-50 (Fig. 2.3, Table 2.1). When the whole protein was incubated with the protease under the same conditions the cleavage was found to be incomplete with most of the protein remaining uncleaved as shown by gel electrophoresis. Two small peptides SP-1-I and SP-1-II (Table 2.1) were recovered after CMC chromatography (4.3.3.2) using 50 mM Na-acetate buffer pH 4.5 and a 0-200 mM NaCl gradient followed by fractionation on Sephadex G-50. In later experiments with $H2B_{(3)}$ Parechinus more complete cleavage was obtained and these results will be discussed in section 2.4.1.1.

2.2.1.4 Chymotryptic digestion

From the chymotryptic digest of the intact protein (4.3.2.3), a number of peptides were developed from a CMC column (Fig.2.4) by a steep salt gradient. The fourth fraction (CH-4) proved to be the expected large basic peptide extending from the amino terminus to the first tyrosine residue. Less basic peptides from the digest were found, mostly in the first fraction but no further purification of these was performed.

2.2.1.5 Thermolysin digestion (4.3.2.4)

Thermolysin generates a complex mixture of peptides. The amino acid compositions and the already established sequences of peptides in the amino terminal part of the protein (CN-I and CN-I, SP-II), made the recovery of a thermolysin peptide starting with valine and containing 6-7 arginines very likely. Peptide CN-I was therefore digested with thermolysin (Boehringer) and a peptide (CN-I, Th-III-4) was recovered after Sephadex G-50 and CMC chromatography with valine

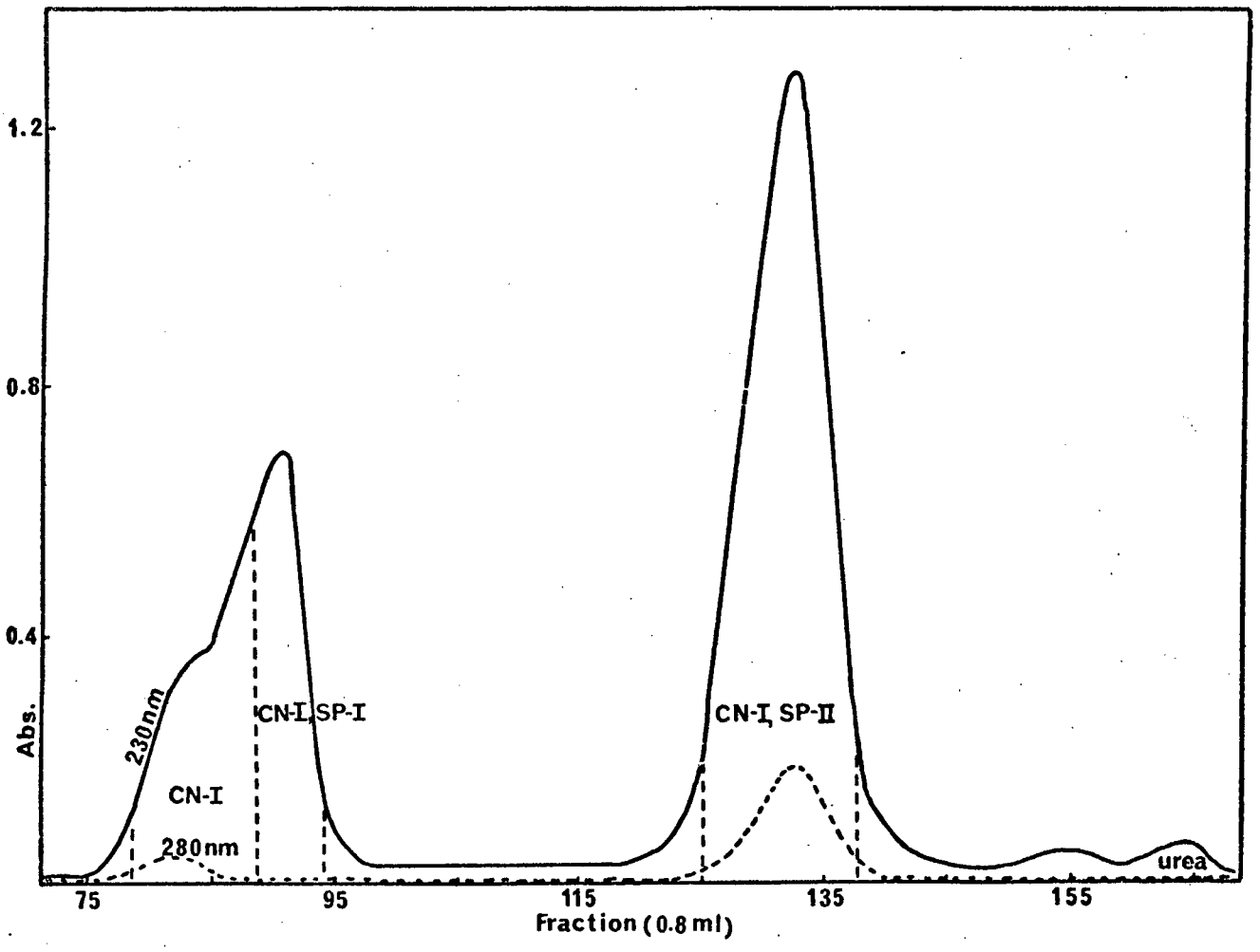


Fig. 2.3: Fractionation of S.aureus protease digested CN-I of $H2B_{(2)}$ Parechinus Column 10 x 2000 mm Sephadex G-50 fine, eluant: 10 mM HCl, cuvette:10 mm, sample:450 n moles applied in 0.5 ml 8 M urea in eluant, yield:about 400 n moles CN-I, SP-II.

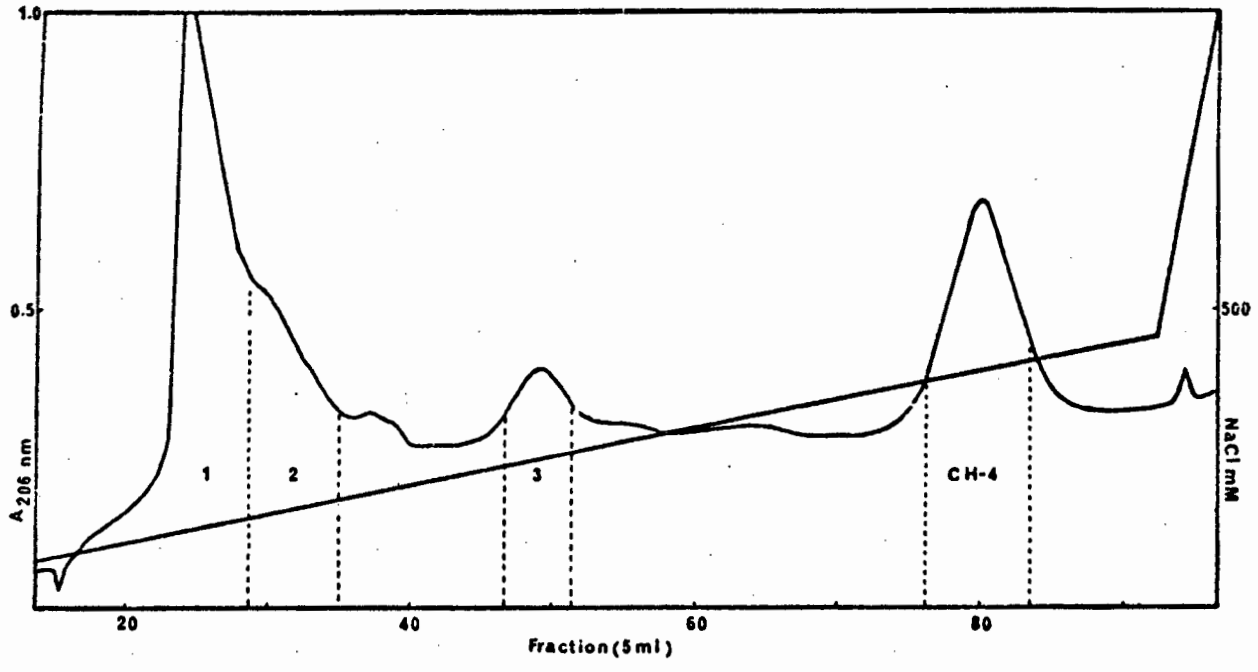


Fig. 2.4: Fractionation of a chymotryptic digestion of $H2B_{(2)}$ Parechinus Column: 16 x 300 mm CMC, eluant: 50 mM Na-acetate/HCl pH 4.5, linear gradient: 0-600 mM NaCl (total volume 600 ml), flow rate: 30 ml/hour, cuvette: 3 mm, sample 40 mg applied in 2 ml 8 M urea in eluant. Yield CH-4 8 mg.

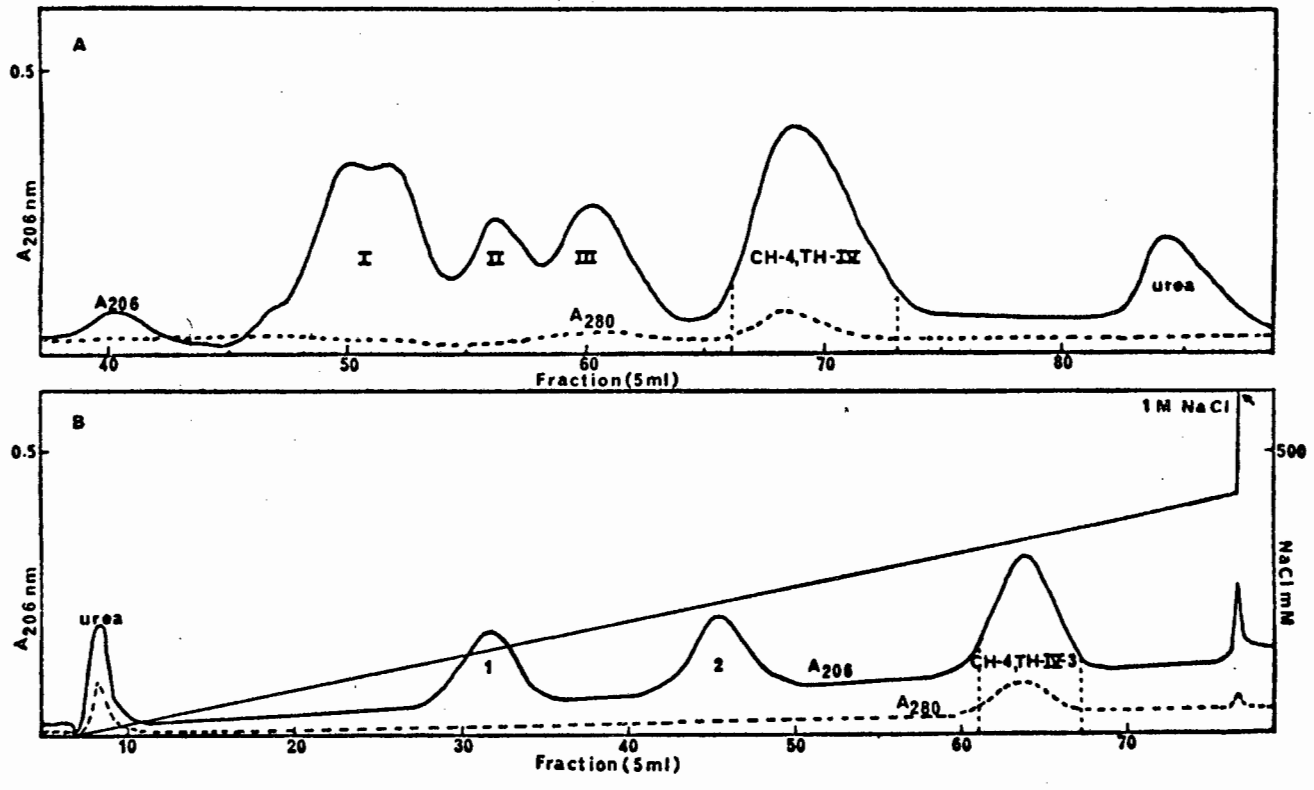


Fig. 2.5: Fractionation of a thermolysin digestion of CH-4 of $H2B_{(2)}$ Parechinus.

A. Column:26 x 1000 mm Sephadex G-50 medium, eluant:10 mM HCl, cuvette 3 mm, sample 15 mg thermolysin digested CH-4 (Fig. 2.4).

B. Column:16 x 300 mm CMC, eluant:50 mM Na-acetate/HCl pH 4.5, linear gradient:0-600 mM NaCl (total volume 500 ml), flow rate: 30 ml/hour, cuvette:3 mm, sample:fraction IV from A (above) yield:1.0 mg CH-4, TH-IV-3.

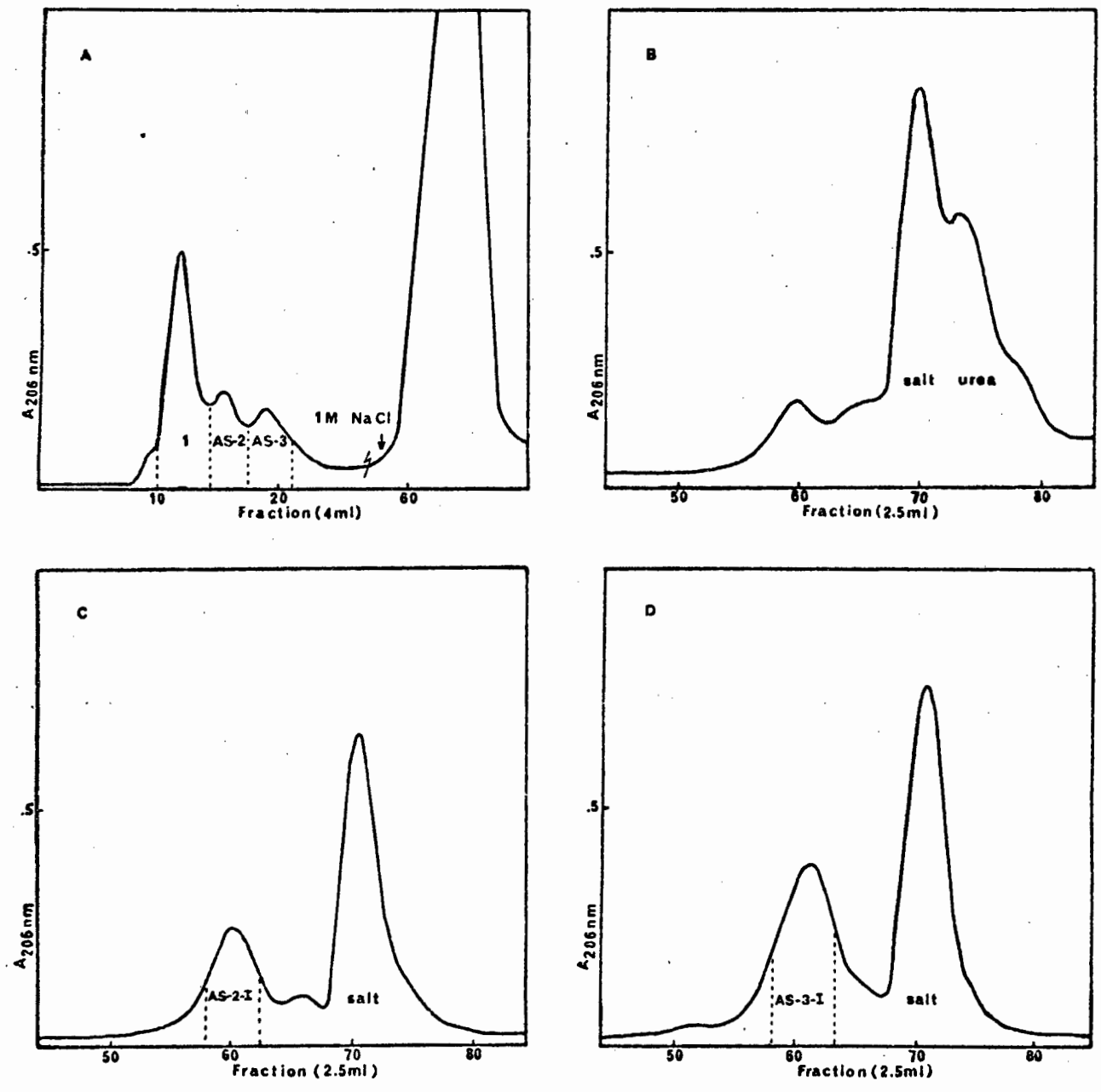


Fig. 2.6: Fractionation of dilute acid hydrolyzed H2B (2) Parechinus.

A. Column: 16 x 300 mm CMC, eluant: 50 mM Na-acetate/HCl pH 4.5, flow rate: 25 ml/hour, cuvette: 3 mm, sample: 30 mg in 1 ml 8 M urea in eluant.

B - D. Column: 16 x 1000 cm Sephadex G-50 medium, eluant: 10 mM HCl.

B. Sample: fraction 1 of A - note most of 206 nm absorption is due to urea.

C. Sample: fraction 2 of A.

D. Sample: fraction 3 of A.

as the N-terminus. But apparently it was heterogeneous at the C-terminus due to incomplete cleavage at the closely spaced isoleucine residues at positions 57 and 59. This peptide gave a single sequence for 13 steps but did not have integral values for tyrosine and isoleucine when the amino acid composition was examined. Another similar peptide was obtained after 4½ hours of digestion with thermolysin of CH-4 (Fig. 2.5A & B). This thermolysin peptide (CH-4, TH-IV-3) had the expected amino acid composition (Table 2.1) for a pure peptide extending from residues 44-55.

2.2.1.6 Cleavage in dilute acid

The mixture of peptides resulting from the cleavage of whole protein with dilute acid (4.3.1.2) was applied to a CMC column (4.3.3.2). From the acidic or neutral peptides which eluted with equilibration buffer, two peptides containing methionine could be recovered (Fig.2.6). Both peptides had the same N-terminal sequence but their amino acid compositions indicated cleavage after Asn-81 in the case of AS-2-I and cleavage before Asp-86 in AS-3-I. AS-2-I must have been produced by cleavage of a Asn-Ser bond while AS-3-I represents the expected cleavage at Asp with loss of that residue from the peptide mixture (free aspartic acid generated).

2.2.2 Alignment of peptides

Histone $H2B_{(2)}$ *Parechinus* was cleaved by chemical and enzymatic means into eight overlapping peptides (Fig. 2.7). The amino acid compositions of the uncleaved protein and of the overlapping peptides (Table 2.1) correspond to the composition found by sequence analysis (Table 2.2). The amino acid sequence of sperm histone $H2B_{(2)}$ *Parechinus* obtained from these data is given (Fig. 2.8).

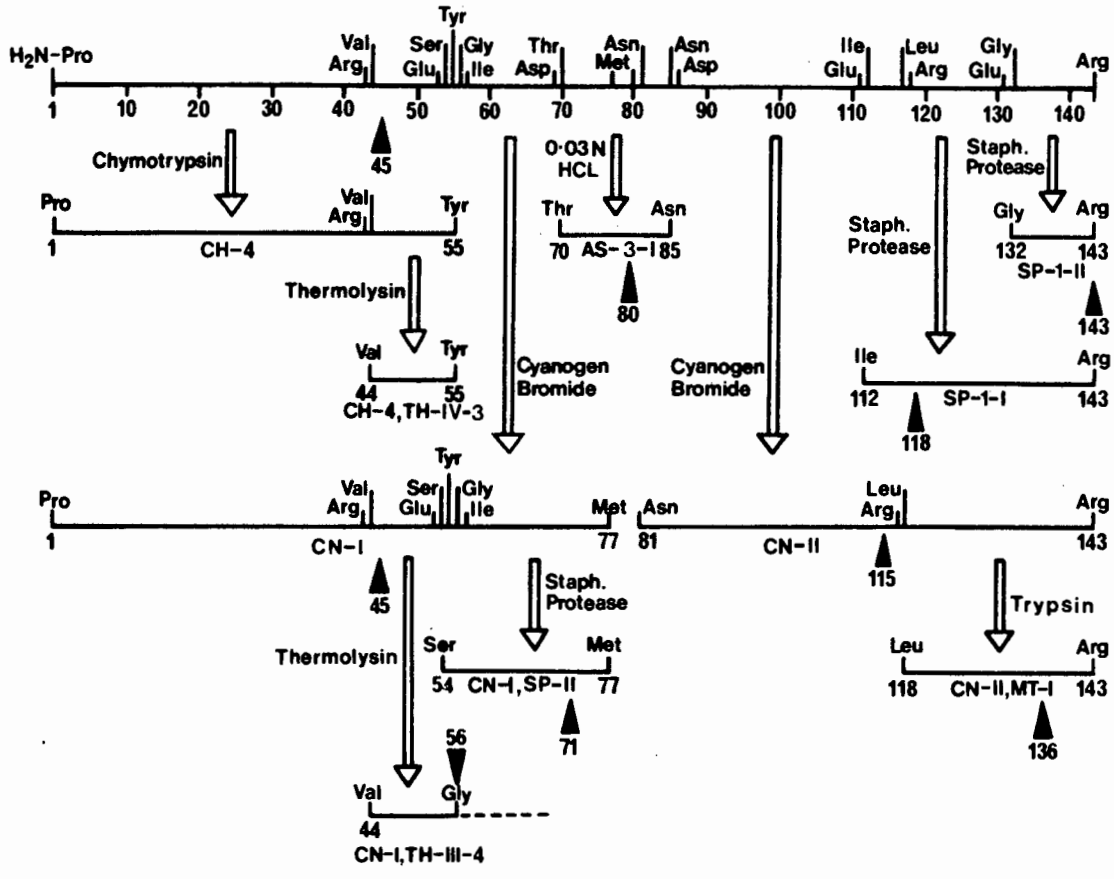


Fig. 2.7 *Alignments of peptides generated from sperm histone* $H2B_{(2)\text{Parechinus}}$. Arrow with number indicates final residue sequenced in that peptide

Reproduced from Strickland, W.N. (1976)

Table 2.1 *Amino acid compositions (number of residues) of* $H2B_{(2)\text{Parechinus}}$ *and of peptides used in the sequence analysis*
Figures in parentheses are those found from the sequence analysis.

Amino acid	$H2B_{(2)}$ (1–143)	CN-I (1–77)	CH-4, TH-IV-3 (44–55)	CN-I SP-II (54–77)	AS-3-I (70–85)	CN-II (81–143)	SP-1-I (112–143)	CN-II, MT-I (118–143)	SP-1-II (132–143)
Aspartic acid	4.79 (5)	1.13 (1)	0	1.00 (1)	2.00 (2)	3.68 (4)	0	0	0
Threonine	7.77 (9)	2.26 (2)	0	0.75 (1)	1.21 (1)	6.76 (7)	4.77 (5)	3.56 (4)	3.76 (4)
Serine	15.72 (18)	8.14 (9)	0.72 (1)	2.23 (3)	3.79 (4)	7.19 (8)	1.74 (2)	1.65 (2)	1.00 (1)
Glutamic acid	8.33 (8)	2.41 (2)	1.02 (1)	1.44 (1)	0.52 (0)	6.21 (6)	2.93 (3)	2.20 (2)	0
Proline	7.70 (8)	7.13 (7)	0	1.14 (1)	0	0.93 (1)	0.79 (1)	0.92 (1)	0
Glycine	13.14 (13)	9.88 (10)	0.22	1.98 (2)	1.24 (1)	2.97 (3)	1.85 (2)	2.32 (2)	1.12 (1)
Alanine	11.83 (12)	5.22 (5)	0.08	1.11 (1)	1.66 (1)	7.25 (7)	3.97 (4)	3.02 (3)	1.00 (1)
Valine	10.60 (11)	4.29 (4)	1.41 (2)	2.05 (2)	2.21 (2)	6.10 (6)	2.95 (3)	2.15 (2)	0.99 (1)
Methionine	2.02 (2)	–[a]	0	0.34 (1)	2.09 (2)	0	0	0	0
Isoleucine	4.80 (5)	2.94 (3)	0.09	2.64 (3)	1.27 (1)	1.70 (2)	0.79 (1)	0	0
Leucine	6.08 (6)	1.15 (1)	0	1.00 (1)	0.0	4.67 (5)	3.95 (4)	3.76 (4)	0
Tyrosine	4.13 (4)	2.98 (3)	1.07 (1)	2.80 (3)	0.0	0.93 (1)	0.89 (1)	0.94 (1)	0.98 (1)
Phenylalanine	2.18 (2)	0	0	0	1.09 (1)	1.76 (2)	0	0	0
Lysine	14.07 (14)	11.19 (11)	1.14 (1)	2.18 (2)	0	2.64 (3)	3.06 (3)	3.02 (3)	2.02 (2)
Histidine	1.80 (2)	0.97 (1)	0	1.00 (1)	0	0.72 (1)	0.89 (1)	0.97 (1)	0
Arginine	24.78 (24)	17.24 (17)	6.38 (6)	1.30 (1)	1.43 (1)	7.09 (7)	2.04 (2)	1.16 (1)	1.33 (1)

[a] Detected as homoserine lactone.

Reproduced from Strickland, W.N. (1976).

2.2.3 Sequence analysis of H2B$_{(2)}$ *Parechinus*

Using peptide CN-I and uncleaved protein in different runs, assignment of residues 1-45 could be made via automated sequence analysis (Table 2.2A). Peptide CN-I was subsequently cleaved into two large peptides by *Staphylococcus aureus* protease. One of these (CN-I, SP-II)was used to identify residues 54-71 (Table 2.2B). Two peptides were isolated namely, CH-4, TH-IV-3 and CN-I, TH-III-4 bridging the region from residue 44 into the N-terminal end of CN-I, SP-II. CH-4, TH-IV-3 on amino acid analysis corresponding to residues 44-55 (Table 2.1), however, did not sequence well on the automatic sequencer. CN-I, TH-III-4 beginning at residue 44 and apparently heterogeneous at the C-terminus was sequenced to identify residues 44-56 (Table 2.2C).

The total amino acid composition of H2B$_{(2)}$ *Parechinus* indicates the presence of two methionine residues. Therefore, cleavage by cyanogen bromide should have yielded three peptides but only two (CN-I and CN-II) were recovered. However, from a mixture of peptides produced by dilute acid hydrolysis peptide AS-3-I was isolated containing two methionine residues. In this peptide residues 70-80 were identified (Table 2.2D). This peptide bridged the gap due to the missing cyanogen bromide peptide.

The second cyanogen bromide peptide CN-II was used to sequence residues 81-115. (Table 2.2E). To position the remaining 28 C-terminal amino acids two peptides which were isolated from a mixture of peptides produced by cleavage of the whole protein or peptide CN-II by *Staphylococcus aureus* protease were useful. The two peptides used in the sequence analysis were SP-1-I (residues 112-143) and SP-1-II (residues 132-143). These two peptides overlap. Residues 112-118 were identified by sequence analysis of SP-1-I (Table 2.2F) while residues 132-142 were identified by sequence analysis of SP-1-II. (Table 2.2H).

Table 2.2: Automatic sequencing of histone $H2B_{(2)}$ *Parechinus* and its peptides.

The amount of peptide subjected to degradation varied between 0.7 and 7 mg. Amounts of amino acid derivatives below 1 n mole have been recorded as 0. (SPITC) following peptide identification symbol = 4-sulfophenylisothiocyanate-reacted peptide. (Gly) following peptide identification symbol = aspartyl and glutamyl glycine methyl ester of the peptide. R = yield of amino acid derivative assigned to that position, R-1 = yield of that amino acid in the preceding cycle, and R + 1 in the following cycle. * = gas chromatographic identification; — = amino acid analysis after hydrolysis; ( ) = sequence position; - = not quantified; a = on hydrolysis accompanied by an equimolar rise in glycine; b = determined in a separate run; c = determined as dansyl derivative, where glutamic or aspartic amides occupy the N-terminus only the parent acid can be identified; d = overlap with adjacent peptides; e = determined with carboxypeptidase. Amounts given for asparagine and glutamine if determined via gas chromatography represent the sum of the corresponding amino acid and its amide.

Table 2.2 A

Peptide :		Intact protein			Peptide :		Intact protein		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1	Pro[c]	-	300*	30	24	Ser[b]	74	148*	98
2	Arg		223	0	25	Pro[b]	8	47*	37
3	Ser[b]	30	200*	0	26	Lys	22	47*	45
4	Pro	73	190*	24	27	Arg	29	52	58
5	Ala	17	268*	83	28	Gly	49	70*	90
6	Lys	24	145*	-	29	Gly	70	90*	86
7	Thr	0	173*	15	30	Lys	25	58	32
8	Ser[b]	33	188*	1	31	Gly	86	124*	97
9	Pro	15	149*	52	32	Ala[b]	10	37*	29
10	Arg	25	105	81	33	Lys[b]	39	60*	46
11	Lys	25	128	62	34	Arg[b]	39	71	54
12	Gly	50	170*	84	35	Ala[b]	20	34*	23
13	Ser[b]	52	143*	64	36	Gly[b]	28	51*	49
14	Pro	0	73*	47	37	Lys[b]	30	46*	43
15	Arg	24	79	71	38	Gly[b]	40	64*	59
16	Lys	21	73	64	39	Gly[b]	64	59*	46
17	Gly	36	80*	96	40	Arg[b]	56	70	79
18	Ser[b]	49	188*	64	41	Arg[b]	70	79	93
19	Pro	0	44*	33	42	Arg[b]	79	93	104
20	Ser[b]	64	113*	75	43	Arg[b]	93	104	70
21	Arg	25	70	64	44	Val[b,d]	12	20*	33
22	Lys	22	56	57	45	Val[b,d]	20	33*	29
23	Ala	33	81*	64					

Table 2.2 B

Peptide :		CN-1, TH-III-4			Peptide :		CN-1, TH-III-4		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (44)	Val[d,c]	0	380*	380	8 (51)	Arg	438	400	431
2 (45)	Val[d]	380	380*	103	9 (52)	Arg	400	431	296
3 (46)	Lys	140	350*	80	10 (53)	Glu	5	151*	151
4 (47)	Arg	22	250	380	11 (54)	Ser[d]	-	-	-
5 (48)	Arg	250	380	430	12 (55)	Tyr[d]	4	20*	27
6 (49)	Arg	380	430	438	13 (56)	Gly[d]	7	26*	45
7 (50)	Arg	430	438	400					

Table 2.2 C

Peptide :		CN-I, SP-II			Peptide :		CN-I, SP-II		
			nmol					nmol	
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (54)	Ser[d,c]		360*	120	10 (63)	Leu	19	342*	114
2 (55)	Tyr[d]	0	725*	61	11 (64)	Lys	37	294*	120
3 (56)	Gly[d]	11	734*	122	12 (65)	Gln	3	355*	64
4 (57)	Ile	11	456*	94	13 (66)	Val	10	155*	62
5 (58)	Tyr	8	360*	154	14 (67)	His	10	135	130
6 (59)	Ile	76	383*	135	15 (68)	Pro	3	75*	47
7 (60)	Tyr	153	389*	103	16 (69)	Asp	0	128*	144
8 (61)	Lys	13	354*	-	17 (70)	Thr[d]	96	190*	173
9 (62)	Val	3	432*	113	18 (71)	Gly[d]	11	196*	-

Table 2.2 D

Peptide :		AS-3-I			Peptide :		AS-3-I		
			nmol					nmol	
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (70)	Thr[d,c]	-	135*	8	7 (76)	Ala	4	70*	4
2 (71)	Gly[d]	45	140*	18	8 (77)	Met	3	52*	20
3 (72)	Ile	4	130*	21	9 (78)	Ser	21	49*	28
4 (73)	Ser	0	56*	65	10 (79)	Val	2	18*	7
5 (74)	Ser	56	56*	20	11 (80)	Met	5	10*	-
6 (75)	Arg	15	91	10					

Table 2.2 E

Peptide:		CN-II(SPITC; Gly)			Peptide:		CN-II(SPITC; Gly)		
			nmol					nmol	
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (81)	Asn[b,c]	-	556*	43	19 (99)	Thr	0	96*	34
2 (82)	Ser	0	262*	0	20 (100)	Ser	0	145*	0
3 (83)	Phe	0	298*	0	21 (101)	Ala	0	99*	40
4 (84)	Val	0	165*	0	22 (102)	Asn	0	100*	0
5 (85)	Asn	0	167*	139	23 (103)	Arg	5	38	53
6 (86)	Asp[a]	167	380*	-	24 (104)	Arg	38	53	-
7 (87)	Val	0	158*	14	25 (105)	Ser	0	101*	60
8 (88)	Phe	0	260*	24	26 (106)	Thr	0	79*	31
9 (89)	Glu[a]	0	250*	-	27 (107)	Val	0	53*	25
10 (90)	Arg	0	57	-	28 (108)	Ser	0	93*	113
11 (91)	Ile	0	260*	33	29 (109)	Ser	93	113*	0
12 (92)	Ala	0	210*	36	30 (110)	Arg	4	25	-
13 (93)	Gly	0	294*	0	31 (111)	Glu[a]	2	28	23
14 (94)	Glu[a]	0	200	0	32 (112)	Ile[d]	7	33*	18
15 (95)	Ala	0	188*	0	33 (113)	Gln[d]	-	-	-
16 (96)	Ser	0	209*	0	34 (114)	Thr[d]	13	39*	20
17 (97)	Arg	0	60	-	35 (115)	Ala[d]	10	21*	-
18 (98)	Leu	0	127*	41					

Table **2.2 F**

Peptide :		SP-1-I			Peptide :		SP-1-I		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (112)	Ile[d,c]	0	133*	0	5 (116)	Val	0	112*	21
2 (113)	Gln[d]	0	92*	0	6 (117)	Arg	0	30	-
3 (114)	Thr[d]	0	114*	0	7 (118)	Leu[d]	0	57*	67
4 (115)	Ala[d]	0	73*	0					

Table **2.2 G**

Peptide :		CN-II-MT-I			Peptide :		CN-II-MT-I		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (118)	Leu[d,c]		630*	870	10 (127)	His	0	79	0
2 (119)	Leu	632	870*	643	11 (128)	Ala	56	292*	110
3 (120)	Leu	870	643*	71	12 (129)	Val	0	103*	104
4 (121)	Pro	0	506*	91	13 (130)	Ser	0	80*	26
5 (122)	Gly	0	879*	250	14 (131)	Glu	0	38*	48
6 (123)	Glu	0	349*	123	15 (132)	Gly[d]	0	79*	-
7 (124)	Leu	0	220*	73	16 (133)	Thr[d]	-	-	-
8 (125)	Ala	0	204*	35	17 (134)	Lys[d]	-	-	-
9 (126)	Lys	0	101	48	18 (135)	Ala[d]	0	63*	-
					19 (136)	Val[d]	22	38*	

Table **2.2 H**

Peptide:		SP-1-II			Peptide:		SP-1-II		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (132)	Gly[d,c]	-	110*	0	7 (138)	Lys	0	12	-
2 (133)	Thr[d]	0	60*	49	8 (139)	Tyr	0	48*	-
3 (134)	Lys[d]	0	16*	-	9 (140)	Thr	7	26*	33
4 (135)	Ala[d]	0	44*	39	10 (141)	Thr[e]	26	33*	28
5 (136)	Val[d]	0	46*	38	11 (142)	Ser[e]	0	12*	0
6 (137)	Thr	0	49*	23	12 (143)	Arg[e]	-	-	-

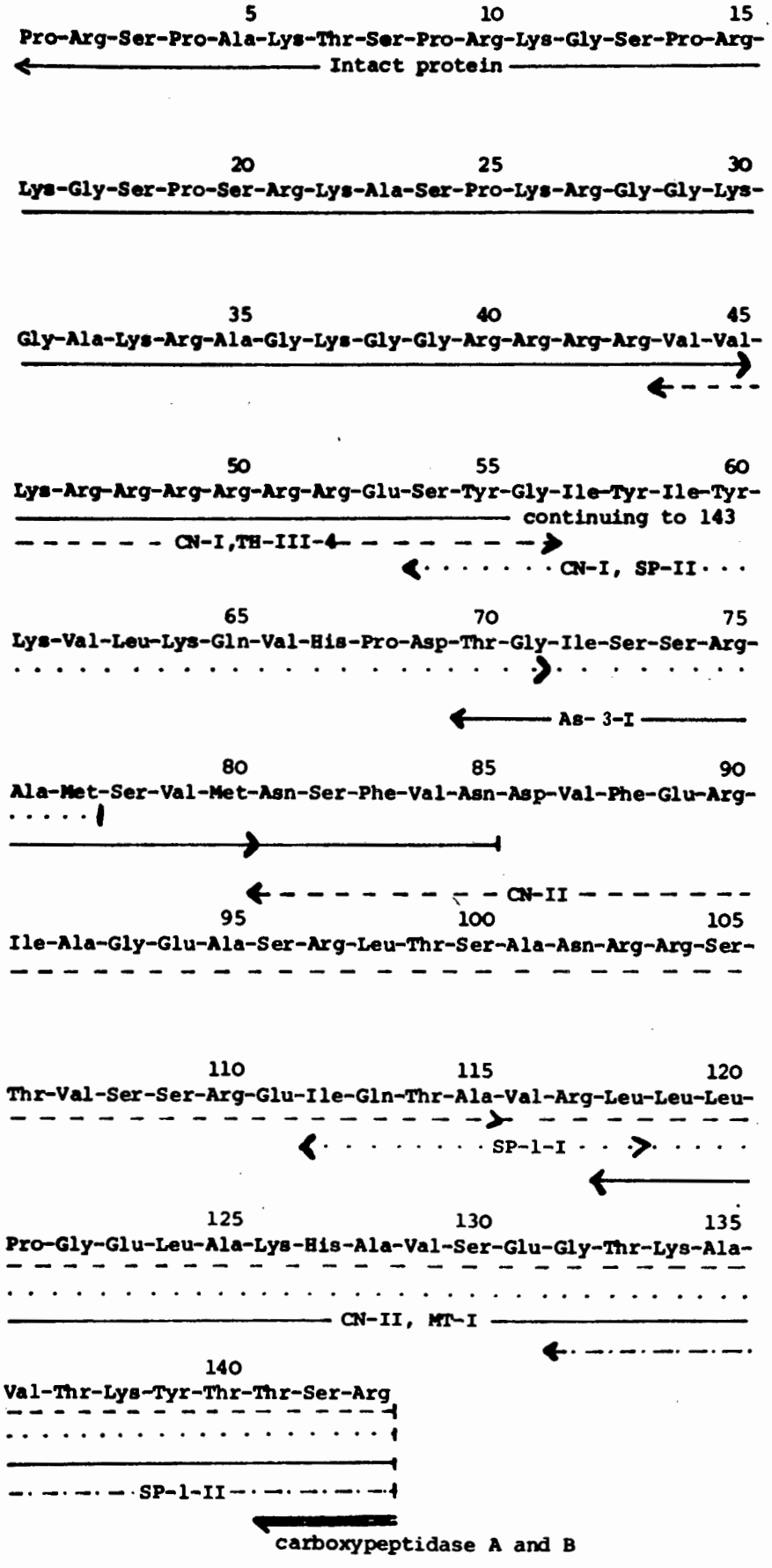


Fig. 2.8: Primary structure of sperm histone H2B(2) Parechinus.

This peptide bridged the gaps between the two sequences established in the two previous peptides. The C-terminal peptide CN-II, MT-I which was produced by tryptic digestion of maleylated CN-II was sequenced from residue 118-136 (Table 2.2G). Residues 134 and 135 could not be positively identified in peptide CN-II, MT-I but were positively identified in peptide SP-1-II.

The last three residues of the protein sequence (Thr-Ser-Arg-COOH) were also identified by analysis of the free amino acids produced by sequential digestion of the uncleaved protein with a mixture of carboxypeptidases A and B (Fig.1.10).

2.3 SEQUENCE OF $H2B_{(1)}$ PARECHINUS

2.3.1 Generation and purification of peptides

All peptides used to sequence $H2B_{(1)}$ were generated and purified essentially as described in section 2.2.1 for $H2B_{(2)}$ with the exception of CH-3. $H2B_{(1)}$ has one more tyrosine than $H2B_{(2)}$ (Table 1.2) and cleavage at this residue permitted isolation of a peptide beginning at residue 103. Fig. 2.9 illustrates the chromatography of the chymotryptic peptides from $H2B_{(1)}$.

2.3.2 Alignment of peptides

The overlapping peptides used in the elucidation of the $H2B_{(1)}$ Parechinus sequence are shown in Fig. 2.10. The entire sequence of 144 residues has been elucidated by sequencing the uncleaved protein and eight peptides. The amino acid composition of the uncleaved protein and of the eight peptides are in good agreement with the sequence found. (Table 2.3, Fig.2.11).

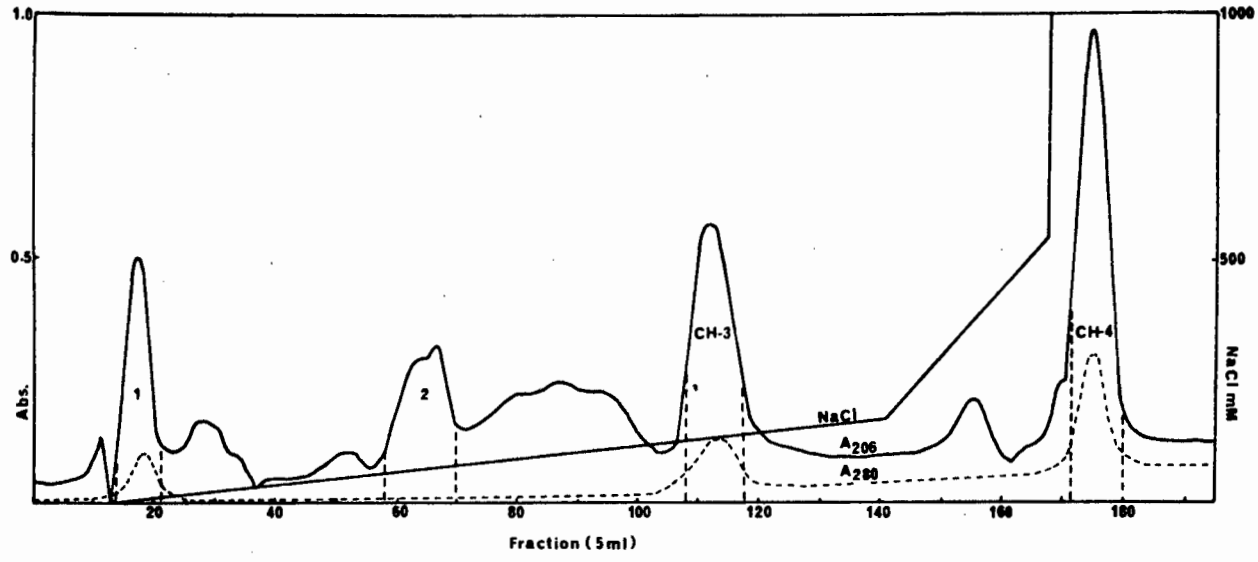


Fig. 2.9: Fractionation of a chymotryptic digestion of $H2B_{(1)}$ Parechinus. Column: 16 x 350 mm CMC, eluant: 50 mM Na-acetate/HCl pH 4.5, NaCl gradients as indicated, flow rate: 30 ml/hour, cuvette: 3 mm, sample: 32 mg in 2 ml 8 M urea in eluant.

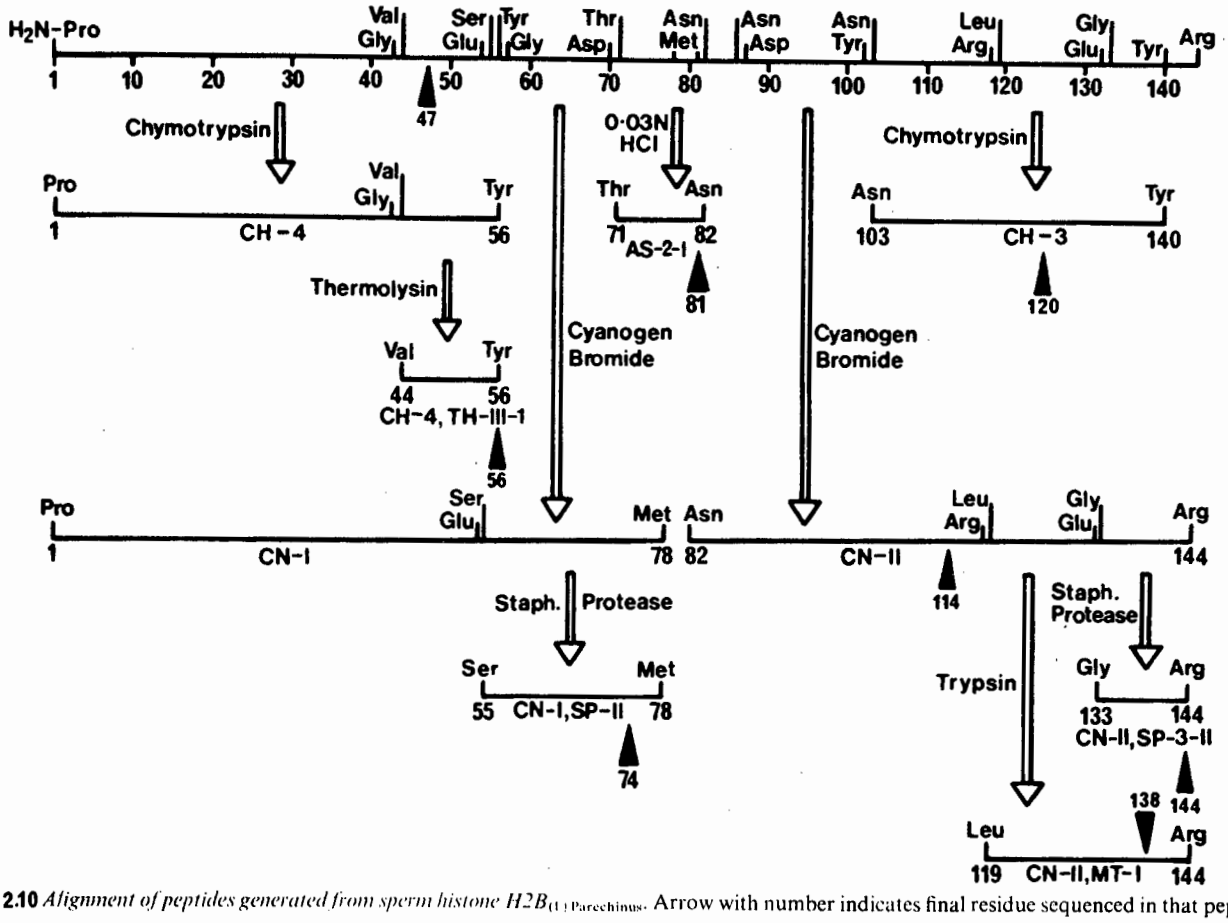


Fig. 2.10 *Alignment of peptides generated from sperm histone* $H2B_{(1)\,Parechinus}$. Arrow with number indicates final residue sequenced in that peptide

Table 2.3 *Amino acid compositions (number of residues) of* $H2B_{(1)\,Parechinus}$ *and of peptides used in the sequence analysis*

Amino acid	$H2B_{(1)}$ (1–144)	CN-I (1–78)	CH-4, TH-III-1 (44–56)	CN-I, SP-II (55–78)	AS-1-I (71–82)	CN-II (82–144)	CH-3 (103–140)	CN-II, MT-I (119–144)	CN-II, SP-3-II (133–144)
Aspartic acid	4.65 (5)[a]	1.10 (1)	0	0.91 (1)	0.70 (1)	3.82 (4)	0.87 (1)	0	0
Threonine[b]	11.15 (12)	3.85 (4)	0	0.83 (1)	0.74 (1)	7.18 (8)	3.52 (4)	3.31 (4)	3.57 (4)
Serine[b]	13.33 (15)	6.95 (8)	0.70 (1)	2.77 (3)	2.83 (3)	4.62 (6)	3.37 (4)	1.75 (2)	1.01 (1)
Glutamic acid	11.15 (11)	4.97 (5)	1.70 (2)	1.03 (1)	0.30 (0)	6.49 (6)	4.16 (4)	2.04 (2)	0
Proline	6.05 (7)	5.74 (6)	0	1.00 (1)	0.00 (0)	1.60 (1)	1.02 (1)	1.04 (1)	0
Glycine	14.62 (14)	10.19 (11)	0.30 (0)	1.91 (2)	1.26 (1)	3.16 (3)	5.46 (2)[d]	2.18 (2)	0.96 (1)
Alanine	10.08 (10)	3.33 (3)	0	0.93 (1)	1.62 (1)	6.80 (7)	4.04 (4)	3.14 (3)	0.96 (1)
Valine	11.58 (12)	4.44 (4)	1.82 (2)	1.93 (2)	1.26 (1)	6.68 (7)	4.95 (5)	2.26 (2)	0.98 (1)
Methionine	1.67 (2)	–[c]	0	–[c]	1.57 (2)	0	0	0	0
Isoleucine[b]	3.54 (4)	2.81 (3)	0	3.01 (3)	0.86 (1)	1.01 (1)	0	0	0
Leucine	5.84 (6)	1.24 (1)	0	0.90 (1)	0	4.84 (5)	4.01 (4)	3.96 (4)	0
Tyrosine[b]	4.37 (5)	2.76 (3)	0.85 (1)	2.82 (3)	0	1.79 (2)	0.80 (1)	1.08 (1)	0.93 (1)
Phenylalanine	1.57 (2)	0	0	0	0	1.85 (2)	0	0	0
Lysine	15.80 (16)	12.19 (13)	0.88 (1)	2.44 (2)	0.17 (0)	2.84 (3)	3.19 (3)	2.82 (3)	1.92 (2)
Histidine	1.69 (2)	0.84 (1)	0	0.90 (1)	0	1.00 (1)	0.85 (1)	0.96 (1)	0
Arginine	22.83 (21)	13.61 (14)	6.39 (6)	1.32 (1)	1.34 (1)	6.82 (7)	4.05 (4)	1.00 (1)	1.30 (1)

[a] Figures in parentheses are residues found from the sequence analysis.
[b] Not corrected for losses during and incompleteness of hydrolysis.
[c] Detected as homoserine lactone.
[d] Three glycine residues covalently bonded to γ-carboxyl groups of glutamic acid included.

Table 2.4: Automatic sequencing of histone H2B(1) Parechinus and its peptides

The amount of peptide subjected to degradation varied between 0.7 and 7 mg. Amounts of amino acid derivatives below 1 n mole have been recorded as 0. (SPITC) following peptide identification symbol = 4-sulfophenylisothiocyanate-reacted peptide. (Gly) following peptide identification symbol = aspartyl and glutamyl glycine methyl ester of the peptide. R = yield of amino acid derivative assigned to that position, R-1 = yield of that amino acid in the preceding cycle, and R + 1 in the following cycle. * = gas chromatographic identification; — = amino acid analysis after hydrolysis; ( ) = sequence position; - = not quantified; a = on hydrolysis accompanied by an equimolar rise in glycine; b = determined in a separate run; c = determined as dansyl derivative, where glutamic or aspartic amides occupy the N-terminus only the parent acid can be identified; d = overlap with adjacent peptides; e = determined with carboxypeptidase. Amounts given for asparagine and glutamine if determined via gas chromatography represent the sum of the corresponding amino acid and its amide.

Table 2.4A

Peptide :	Intact protein				Peptide :	Intact protein			
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1	Pro c	-	158*	30	24	Gly	43	71*	92
2	Ser b	46	224*	33	25	Gly	71	92*	82
3	Gln	0	309*	47	26	Lys	30	45	54
4	Lys	15	178	42	27	Gly	82	115*	139
5	Ser b	0	84*	20	28	Gly	115	139*	111
6	Pro	0	122*	38	29	Lys	25	51	42
7	Thr	0	300*	41	30	Gly b	51	69*	43
8	Lys	22	106	56	31	Ala	14	40*	41
9	Arg	0	139	86	32	Lys	16	32	28
10	Ser b	29	136*	36	33	Arg	19	42	44
11	Pro	0	69*	37	34	Gly	45	69*	89
12	Thr	0	156*	33	35	Gly	69	89*	72
13	Lys	32	85	61	36	Lys b	30	58	32
14	Arg	0	90	73	37	Ala	17	38*	37
15	Ser b	46	211*	51	38	Gly b	32	42*	29
16	Pro	0	73*	34	39	Lys b	22	32	28
17	Thr	20	89*	50	40	Arg	30	44	64
18	Lys	21	62	47	41	Arg	44	64	75
19	Arg	24	59	49	42	Arg	64	75	-
20	Ser	32	97*	39	43	Gly	32	47	63
21	Pro	0	37*	48	44	Val[d]	9	31*	37
22	Gln	10	55*	51	45	Gln[d]	15	40*	37
23	Lys	35	56	22	46	Val[d]	37	38*	38
					47	Lys[d]	12	21	-

Table **2.4 B**

Peptide :		CH-4, TH-III-1			Peptide :		CH-4, TH-III-1		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (44)	Val[d,c]	-	83*	25	8 (51)	Arg	102	106	114
2 (45)	Gln	0	85*	22	9 (52)	Arg	106	114	115
3 (46)	Val	15	48*	-	10 (53)	Arg	114	115	68
4 (47)	Lys	0	43	13	11 (54)	Glu	0	25*	18
5 (48)	Arg	12	52	99	12 (55)	Ser[d]	-	-	-
6 (49)	Arg	52	99	102	13 (56)	Tyr[d]	7	35*	-
7 (50)	Arg	99	102	106					

Table **2.4 C**

Peptide :		CN-1, SP-II			Peptide :		CN-1, SP-II		
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (55)	Ser[d,c]	-	122*	0	11 (65)	Lys	7	80	-
2 (56)	Tyr[d]	9	240*	0	12 (66)	Gln	0	123*	0
3 (57)	Gly	0	200*	0	13 (67)	Val	0	132*	64
4 (58)	Ile	0	254*	52	14 (68)	His	0	35	-
5 (59)	Tyr	0	231*	0	15 (69)	Pro	0	58*	39
6 (60)	Ile	52	274*	59	16 (70)	Asp	0	41*	0
7 (61)	Tyr	0	106*	0	17 (71)	Thr[d]	0	128*	99
8 (62)	Lys	3	112	-	18 (72)	Gly[d]	0	117*	84
9 (63)	Val	0	145*	60	19 (73)	Ile[d]	0	64*	46
10 (64)	Leu	0	169*	55	20 (74)	Ser[d]	0	18*	-

Table 2.4 D

Peptide :		A-2-I			Peptide :		A-2-I		
			nmol					nmol	
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (71)	Thr[d,c]	-	13*	0	7 (77)	Ala	2	7*	6
2 (72)	Gly[d]	0	20*	6	8 (78)	Met	0	6*	2
3 (73)	Ile[d]	6	17*	3	9 (79)	Ser	0	14*	0
4 (74)	Ser[d]	6	14*	14	10 (80)	Val	2	3*	2
5 (75)	Ser	14	21*	0	11 (81)	Met	1	3*	-
6 (76)	Arg	3	17	4					

Table 2.4 E

Peptide :		CN-II (SPITC; Gly)			Peptide :		CN-II (SPITC; Gly)		
			nmol					nmol	
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (82)	Asn[c,b]	-	490*		18 (99)	Leu	0	84*	52
2 (83)	Ser	5	113*	0	19 (100)	Thr	0	65*	93
3 (84)	Phe	4	468*	0	20 (101)	Thr	65	93*	17
4 (85)	Val	0	465*	24	21 (102)	Tyr	0	72*	49
5 (86)	Asn	0	320*	305	22 (103)	Asn[d]	4	38*	-
6 (87)	Asp[a]	320	305	0	23 (104)	Arg[d]	13	33	46
7 (88)	Val	11	294*	52	24 (105)	Arg[d]	33	46	-
8 (89)	Phe	0	424*	26	25 (106)	Ser[d]	0	54*	30
9 (90)	Glu[a]	0	254	0	26 (107)	Thr[d]	37	53*	49
10 (91)	Arg	15	86	-	27 (108)	Val[d]	10	27*	24
11 (92)	Ile	0	274*	59	28 (109)	Ser[d]	15	40*	79
12 (93)	Ala	0	246*	313	29 (110)	Ser[d]	40	79*	29
13 (94)	Ala	246	313*	56	30 (111)	Arg[d]	18	26	-
14 (95)	Glu[a]	12	137	-	31 (112)	Glu[d]	12	21	17
15 (96)	Ala	56	182*	104	32 (113)	Val[d]	20	25*	32
16 (97)	Gly	0	184*	47	33 (114)	Gln[d]	17	39*	-
17 (98)	Arg	13	52	-					

Table 2.4 F

Peptide :	CH-3 (SPITC, Gly)				Peptide :	CH-3 (SPITC, Gly)			
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (103)	Asn[d,c]	-	-		10 (112)	Glu[a,d]	12	128	0
2 (104)	Arg[d]	0	93	139	11 (113)	Val[d]	8	173*	0
3 (105)	Arg[d]	93	139	-	12 (114)	Gln[d]	0	117*	0
4 (106)	Ser[b,d]	0	54*	30	13 (115)	Thr	0	184*	0
5 (107)	Thr[d]	0	173*	0	14 (116)	Ala	0	164*	35
6 (108)	Val[d]	0	155*	127	15 (117)	Val	0	153*	37
7 (109)	Ser[d]	0	250*	294	16 (118)	Arg	0	41	-
8 (110)	Ser[d]	250	294*	0	17 (119)	Leu[d]	0	139*	161
9 (111)	Arg[d]	5	75	-	18 (120)	Leu[d]	139	161*	-

Table 2.4 G

Peptide :	CN-II, MT-I (Gly)				Peptide :	CN-II, MT-I (Gly)			
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (119)	Leu[d,c]	0	251*	234	11 (129)	Ala	0	80*	54
2 (120)	Leu[d]	251	234*	232	12 (130)	Val	0	59*	42
3 (121)	Leu	234	232*	0	13 (131)	Ser	0	38*	0
4 (122)	Pro	0	112*	0	14 (132)	Glu[a]	0	27	-
5 (123)	Gly	0	297*	0	15 (133)	Gly[d]	0	18*	20
6 (124)	Glu[a]	0	100	0	16 (134)	Thr[d]	0	28*	26
7 (125)	Leu	0	115*	46	17 (135)	Lys[d]	3	7	-
8 (126)	Ala	0	105*	0	18 (136)	Ala[d]	6	21*	21
9 (127)	Lys	0	32	21	19 (137)	Val[d]	5	17*	27
10 (128)	His	0	20	-	20 (138)	Thr[d]	9	33*	32

Table 2.4 H

Peptide:	CN-II,SP-3-II				Peptide:	CN-II,SP-3-II			
		nmol					nmol		
Residue No.	Residue	R-1	R	R+1	Residue No.	Residue	R-1	R	R+1
1 (133)	Gly[d,c]	-	145*	0	7 (139)	Lys	2	46	-
2 (134)	Thr[d]	0	150*	24	8 (140)	Tyr	0	143*	65
3 (135)	Lys[d]	0	91*	-	9 (141)	Thr	0	89*	104
4 (136)	Ala[d]	0	169*	42	10 (142)	Thr[e]	89	104*	48
5 (137)	Val[d]	0	160*	38	11 (143)	Ser[e]	0	24*	0
6 (138)	Thr[d]	0	110*	78	12 (144)	Arg[e]	0	10	-

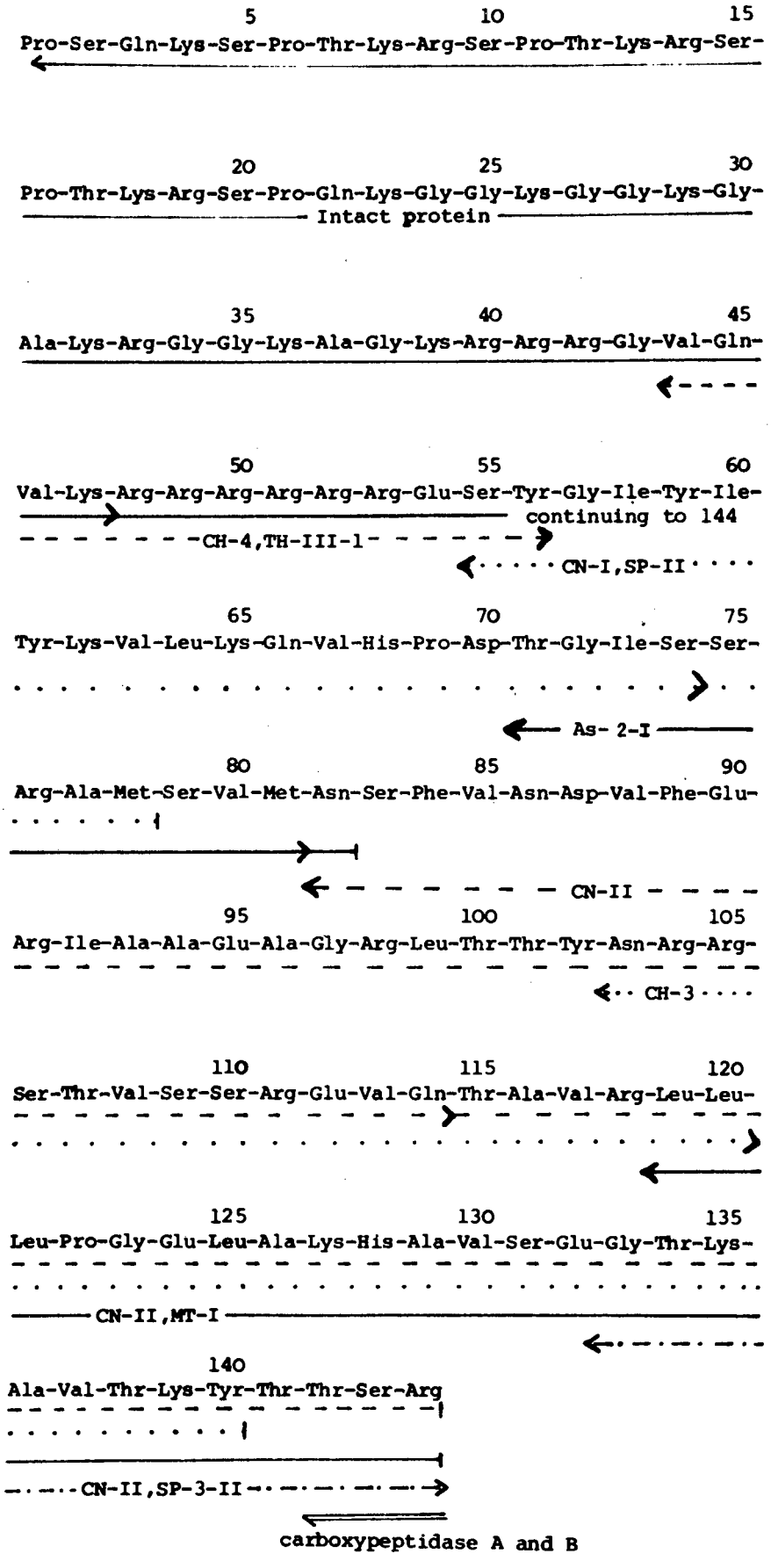


Fig. 2.11: Primary structure of sperm histone H2B (1) Parechinus.

2.3.3 Sequence analysis

In the protein unequivocal assignment of residues 1-45 could be made via automated sequence analysis (Table 2.4A). To entend the sequence towards methionine (residue 78) peptide CH-4, TH-III-1, incorporating the residues 44-56, was subjected to sequence analysis. (Table 2.4B). Though Tyr at position 56 could be clearly identified, the preceding Ser at 55 could only be identified through the analysis of peptide CN-I, SP-II, ranging from residue 55-78 (Table 2.4C). Since the total amino acid composition of $H2B_{(1)}$ indicates two methionine residues, cyanogen bromide cleavage should have yielded 3 peptides; however, only two were recovered. Peptide AS-2-I (residues 71-82) was found (Table 2.4D) to contain the 2 methionines separated by two amino acids as is the case in H2B of calf. Cleavage after asparagine (residue 82) had apparently occurred to yield this peptide.

Resides 82-114 were sequenced in peptide CN-II (Table 2.4E).Peptide CH-3: Its amino acid sequence for positions 103-120 was established (Table 2.4F). Peptide CN-II, MT-I: Residues 119 to 138 were identified (Table 2.4G). Peptide CN-II, SP-3-II: Residues 133 to 144 were identified (Table 2.4H). The C-terminal sequence of the protein Thr-Ser-Arg-COOH was also identified by digestion of the uncleaved protein with a mixture of carboxy-peptidase A and B (Fig. 1.10).

2.4 SEQUENCE OF $H2B_{(3)}$ PARECHINUS

2.4.1 Generation and purification of peptides

Cyanogen bromide and tryptic cleavages and purification of the resulting peptides were done as described for $H2B_{(2)}$ (2.2.1).

2.4.1.1 S. aureus protease

When the whole protein of $H2B_{(3)}$ was subjected to S. aureus protease digestion at pH 4.0 electrophoretic studies showed the cleavage of the protein was much more complete than observed previously with $H2B_{(1)}$ and $H2B_{(2)}$ Parechinus (Fig. 2.12, gel 1). In an experiment to find optimal conditions for cleavage it was noted that the presence of urea seemed to inhibit cleavage at pH 4.0 (Fig. 2.12). Cleavage of $H2B_{(3)}$ by S. aureus protease at pH 7.9 in NH_4HCO_3 and NaH_2PO_4 was also studied by gel electrophoresis (Fig. 2.12, gels 3-6). The presence of urea at this pH apparently did not inhibit cleavage in NH_4HCO_3 and had only a small effect in NaH_2PO_4 buffer. The presence of fast moving faint bands in the cleavages done at basic pH are possibly the results of cleavage at both aspartic and glutamic acid residues which is reported to occur at basic pH (Drapeau, 1976). From these results it can be postulated that the previously noted poor cleavages of $H2B_{(1)}$ and $H2B_{(2)}$ Parechinus (2.2.1.3) were due to the presence of small amounts of urea as a result of incomplete dialysis after CMC purification.

Figure 2.13 illustrates the purification of several SP peptides (see also Table 2.5). No attempt was made to characterize the expected large basic peptide derived from the amino end. Peptides SP-3, SP-5, SP-6 and SP-9 were each sequenced for 10 steps by the DNS-Edman technique. The sequencing results and the amino acid composition of these peptides indicated several interesting features about the CMC separation and the actual cleavage points. Though the elution order from CMC followed the net charge of the peptides, two peptides with the same net charge SP-5 and SP-6 did separate indicating that the amino acid composition and/or the conformation may play a role in separation. Also, there was evidence that at certain glutamic acid residues cleavage was incomplete (Glu-128, SP-5). When the unfractionated

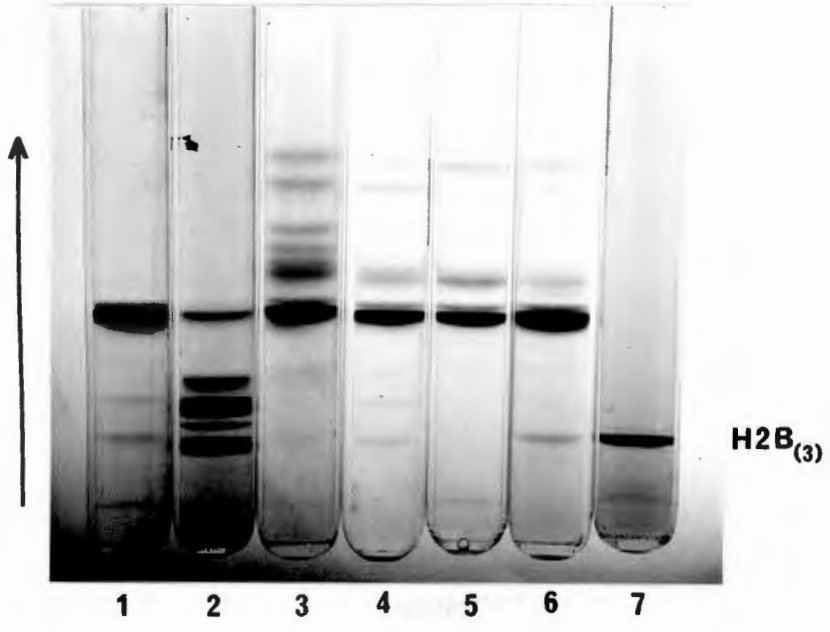


Fig. 2.12: Electrophoretic gels of 18 hour digestions of $H2B_{(3)}$ *Parechinus* with *S. aureus* protease in various buffers. Gels were run for 1 hour and 20 µg of treated or untreated $H2B_{(3)}$ was in each sample.

(1) NH_4 acetate 50 mM pH 4.0

(2) Same buffer as (1) but 2 M in urea

(3) NH_4HCO_3 50 mM pH 7.9

(4) Same buffer as (3) but 2M in urea

(5) NaH_2PO_4 10 mM pH 7.8

(6) Same buffer as (5) but 2 M in urea

(7) Untreated $H2B_{(3)}$

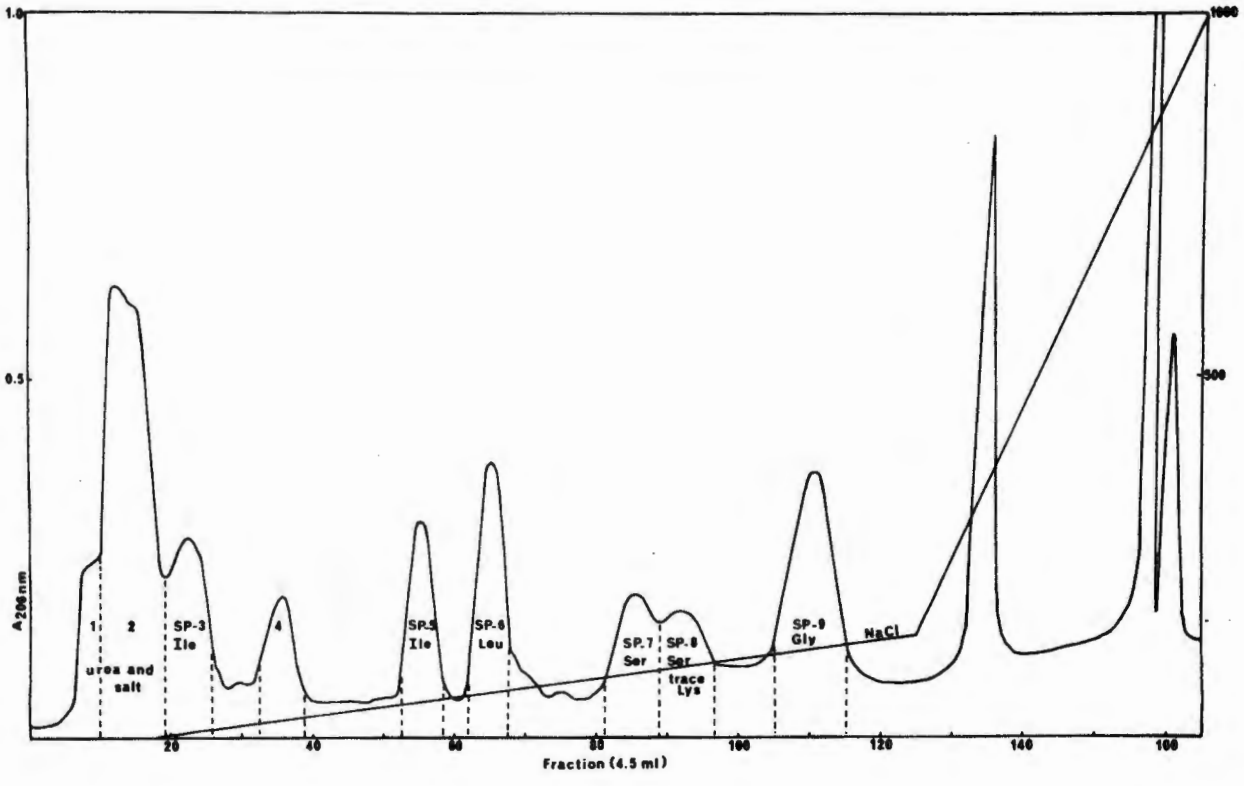


Fig. 2.13: Fractionation of a S.aureus protease digestion of $H2B_{(3)}$ Parechinus. Column: 16 x 300 mm CMC, eluant: 50 mM Na-acetate/HCl pH 4.5, NaCl gradients as indicated, flow rate: 32 ml/hour, cuvette: 3 mm, sample: 29 mg in 5 ml 8 M urea in eluant, yields estimated from amino acid analysis - SP-3: 340 n moles, SP-5: 180 n moles, SP-6: 500 n moles, SP-9: 840 n moles. Amino acids shown are the DNS derivatives found in that fraction.

mixture of peptides was dansylated, DNS-Arg could not be demonstrated as would, however, be expected if cleavage had occurred at residue 94 (Fig.2.16). Peaks 7 and 8 from the CMC fractionation (Fig. 2.13) of the digest both had serine as the major DNS amino group. Also present in these two peaks were 3 yellow spots on the DNS thin layer plates (Fig. 4.1). One of the yellow DNS spots corresponds to the O-tyrosine derivative which is always seen when tyrosine is present in a peptide tested; the other two spots are presumably di- or tri peptides resulting from incomplete hydrolysis of Ile-Tyr bonds of the dansylated peptides. These three yellow DNS derivatives are seen when the whole protein, CN-I, or CN-I,SP-II are dansylated but not in CN-I, SP-I. From this evidence it seems likely that peaks 6 and 7 are peptides resulting from cleavage at the glutamic at residue 58 but have different carboxy-terminals due to different susceptibilities of subsequent Glu-residues. These two fractions were insoluble in the buffers used for desalting and were not recovered in good yield. They could be solubilized in 100% formic acid and stayed in solution when the formic concentration was subsequently reduced to 5%, but did not initially dissolve in 5% formic. These hydrophobic peptides could most likely be purified in good yield from Sephadex columns with 5% formic acid eluants. See also the discussion of a similar tryptic peptide in section 2.2.1.2.

2.4.1.2 Thermolysin digestion

The digestion of CN-I of $H2B_{(3)}$ was for 3 hours with a different batch of thermolysin (Serva 36015) and in the absence of added Ca^{++} ions. Apparently, cleavage was less complete than that seen for $H2B_{(2)}$ (2.2.1.5) as evidenced by the recovery in good yield of CN-I, TH-II-10 which contained an internal Val-Val sequence. Another peptide CN-I, TH-II-9 (Fig. 2.14, Table 2.5) with N-terminal valine and an amino acid composition corresponding to CH-4, TH-IV-3 from $H2B_{(2)}$ (Table 2.1) was also isolated.

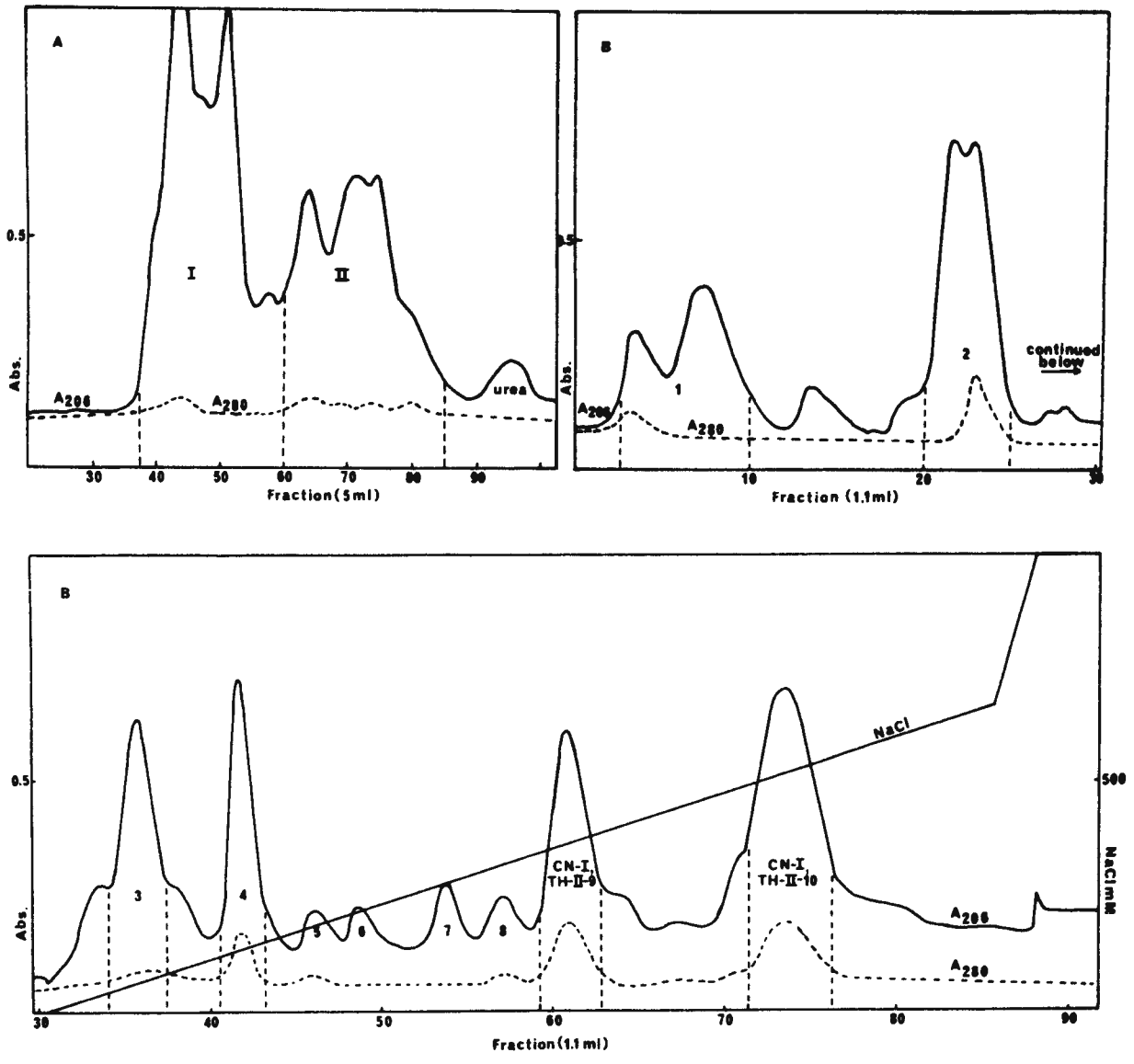


Fig. 2.14: Fractionation of a thermolysin digest of CN-I of $H2B_{(3)}$ Parechinus.

A. Column: 25 x 900 mm Sephadex G-50 fine, eluant: 10 mM HCl, cuvette: 3 mm, sample: 22 mg thermolysin digested CN-I in 1 ml 8 M urea in eluant.

B. Column: 6 x 60 mm, eluant: 50 mM Na-acetate/HCl pH 4.5, NaCl gradients as indicated, flow rate: 6.5 ml/hour, cuvette: 3 mm, yields CN-I, TH-II-9: 100 n moles, CN-I, TH-II-10: 325 n moles.

2.4.2 Alignment of peptides - Sequence analysis

Fig. 2.15 shows the complete sequence of $H2B_{(3)}$ Parechinus and the alignment of peptides used. Amino acid composition of the protein and its peptides are given in Table 2.5. The intact protein was sequenced by the spinning cup method for 42 steps (Table 2.6A). Peptide CN-I, TH-II-10 (residue 40-61) was also sequenced in the spinning cup (Table 2.6B). The residue at step 59 could not be recognized with certainty, but was positively identified as serine on degradation of peptide CN-I, SPII (residue 59-85). The latter peptide was sequenced by a solid state method. Positions at 72 and 80 (4.4.2) were assigned to histidine and arginine respectively by analogy to histone $H2B_{(2)}$ and because the amino acid composition of the peptide includes one histidine and one arginine (Table 2.5). At positions 82 and 85 homoserine was identified but the yields at positions 83 and 84 were too low to make unequivocal assignments. On carboxypeptidase A degradation of CN-I, however, the four following amino acids were identified as the most prominent ones (yields in nmoles) : homoserine 3.8, alanine 2.4, valine 1.9, and serine 0.8. Since carboxypeptidase A does not liberate arginine and releases serine only very slowly (Ambler, 1972) the data are consistent with the assumption that the carboxy-terminal of CN-I is heterogeneous due to incomplete cleavage by cyanogen bromide of a Met-Ser bond (Schroeder et.al., 1969) and that the sequence around the two methionine residues is as in the homologous proteins histone $H2B_{(1)}$ Parechinus and $H2B_{(2)}$ Parechinus -Arg-Ala-Met-Ser-Val-Met- (Fig.3.3).

Residues 86-121 were identified by subjecting CN-II to liquid-state sequencing. (Table 2.6C). Residues 117-128 were identified in the small peptide SP-3 which was sequenced by DNS-Edman degration. Residues 123-147 were identified by solid-state sequencing of peptide MT-1. Position 132 was assigned to histidine on similar grounds as in peptide CN-I, SP-II. The carboxy-terminal amino acid was identified as arginine by digestion of the intact protein with carboxypeptidase A and B (Fig. 1.11).

TABLE 2.5; Amino Acid compositions (number of residues) of $H2B_{(3)}$ *Parechinus* and of peptides used to elucidate its sequence.

	$H2B_{(3)}$	CN-I	CN-I,SP-I	CN-I,SP-II	CN-I,TH-II-9	CN-I,TH-II-10	CN-II	MT-I	SP-3	SP-5	SP-6	SP-9
	1-148	1-85[d]	1-58	59-85[d]	49-61	40-61	86-148	123-148	117-128	117-136	129-136	137-148
Aspartic acid	4.9 (5)	1.1 (1)	0	1.0 (1)	0	0	3.7 (4)	0	0	0	0	0
Threonine	7.4 (9)	1.7 (2)	0.9 (1)	0.9 (1)	0	0	6.0 (7)	3.2 (4)	1.0 (1)	0.9 (1)	0	3.7 (4)
Serine	17.2 (20)	8.7 (11)	5.8 (7)	2.7 (4)	0.8 (1)	0.8 (1)	7.6 (9)	1.8 (2)	0	0.9 (1)	0.9 (1)	1.0 (1)
Glutamic acid	8.2 (8)	2.2 (2)	1.1 (1)	1.0 (1)	1.0 (1)	1.0 (1)	5.9 (6)	2.1 (2)	2.1 (2)	3.0 (3)	1.0 (1)	0
Proline	9.2 (9)	8.5 (8)	6.4 (7)	1.1 (1)	0	0	1.0 (1)	0.9 (1)	1.1 (1)	1.2 (1)	0	0
Glycine	13.0 (13)	10.9 (11)	9.0 (9)	1.9 (2)	0.7 (1)	3.8 (4)	2.4 (2)	2.4 (2)	1.0 (1)	1.0 (1)	0	1.0 (1)
Alanine	12.0 (12)	5.0 (5)	4.0 (4)	1.1 (1)	0	0.9 (1)	7.0 (7)	2.7 (3)	1.0 (1)	3.0 (3)	2.0 (2)	1.0 (1)
Valine	10.4 (11)	4.2 (5)	1.6 (2)	2.0 (2)	1.4 (2)	1.5 (2)	6.0 (6)	1.6 (2)	1.0 (1)	2.0 (2)	1.0 (1)	1.0 (1)
Methionine	1.8 (2)	c (2)[d]	0	c (2)	0	0	0	0	0	0	0	0
Isoleucine	4.8 (5)	2.8 (3)	0	2.7 (3)	0	0	1.9 (2)	0	0.9 (1)	0.9 (1)	0	0
Leucine	6.1 (6)	1.1 (1)	0	1.0 (1)	0	0	4.8 (5)	3.0 (4)	2.9 (3)	3.9 (4)	1.0 (1)	0
Tyrosine	3.9 (4)	2.8 (3)	0	2.7 (3)	1.0 (1)	0.9 (1)	1.0 (1)	0.9 (1)	0	0	0	0.9 (1)
Pheynlalanine	2.0 (2)	0	0	0	0	0	1.9 (2)	0	0	0	0	0
Lysine	15.2 (15)	12.0 (12)	9.9 (10)	1.7 (2)	0.9 (1)	2.0 (2)	3.2 (3)	2.4 (3)	0	1.0 (1)	1.0 (1)	2.0 (2)
Histidine	1.9 (2)	0.9 (1)	0	0.7 (1)	0	0	0.9 (1)	1.0 (1)	0	1.0 (1)	1.0 (1)	0
Arginine	25.2 (25)	17.5 (18)	17.2 (17)	0.9 (1)	5.9 (6)	10.2 (10)	6.8 (7)	1.0 (1)	1.0 (1)	0.9 (1)	0	1.0 (1)
TOTAL:	148	85	58	27	13	22	63	26	12	20	8	12

(a) Figures in parentheses are those found from sequence analysis.

(b) No corrections have been made for losses during hydrolysis or for incomplete hydrolysis.

(c) Detected as homoserine and homoserine lactone but not quantitated.

(d) Incomplete cleavage of Methionine (82) - Serine (83).

TABLE 2.6: - Automatic sequencing of histone $H2B_{(3)}$ Parechinus and its peptides. (SPITC) following peptide identification symbol = 4 - sulfophenylisothiocyanate - reacted peptide. (Gly) following peptide identification symbol = aspartyl and glutamyl glycine methyester of the peptide. R = yield of amino acid derivative assigned to that position, R-1 = yield of that amino acid in the preceding cycle and R+1 in the following cycle.

( ) = sequence position;

a = on hydrolysis accompanied by an equimolar rise in glycine;

b = overlap with adjacent peptides.

Amounts given for asparagine and glutamine represent the sum of the corresponding amino acid and its amide.

TABLE 2.6A: - Identification by HPLC

Intact protein 7.5 mg Protein program

Residue No.	Residue	nano moles			Residue No.	Residue	nano moles		
		R-1	R	R+1			R-1	R	R+1
1	Pro	-	146	20	22	Gly	21	115	62
2	Arg	0	193	25	23	Ser	38	61	55
3	Ser	0	136	27	24	Pro	19	60	39
4	Pro	8	160	29	25	Ser	55	100	62
5	Ala	5	177	33	26	Arg	47	115	112
6	Lys	12	177	38	27	Lys	37	68	51
7	Thr	33	149	6	28	Ala	20	60	50
8	Ser	26	113	40	29	Ser	41	88	61
9	Pro	15	93	44	30	Pro	21	50	26
10	Arg	10	126	82	31	Lys	38	51	34
11	Lys	21	93	61	32	Arg	39	74	65
12	Gly	9	111	67	33	Gly	35	83	119
13	Ser	24	74	52	34	Gly	83	119	53
14	Pro	11	48	38	35	Lys	29	45	35
15	Arg	10	108	45	36	Gly	53	90	54
16	Lys	44	118	75	37	Ala	24	57	36
17	Gly	21	108	61	38	Lys	30	51	36
18	Ser	28	65	46	39	Arg	33	70	46
19	Pro	12	48	31	40	Ala[b]	31	57	29
20	Arg	24	75	56	41	Gly[b]	31	77	54
21	Lys	40	94	70	42	Lys[b]	19	34	27

TABLE 2.6B: - Identification: gas chromatography, and amino acid analysis after acid (underlined yields)

Peptide CNI; TH-II-10 1.7 mg peptide program

Residue No.		Residue	nano moles R-1	nano moles R	nano moles R+1	Residue No.		Residue	nano moles R-1	nano moles R	nano moles R+1
1	(40)	Ala[b]	-	186	109	12	(51)	Lys	5	74	38
2	(41)	Gly[b]		100	65	13	(52)	Arg	30	59	80
3	(42)	Lys[b]	5	93	63	14	(53)	Arg	59	80	104
4	(43)	Gly	65	108	160	15	(54)	Arg	80	104	100
5	(44)	Gly	108	160	115	16	(55)	Arg	104	100	105
6	(45)	Arg	0	33	58	17	(56)	Arg	100	105	111
7	(46)	Arg	33	58	64	18	(57)	Arg	105	111	101
8	(47)	Arg	58	64	81	19	(58)	Glu	0	17	27
9	(48)	Arg	64	81	55	20	(59)	(Ser)[b]			
10	(49)	Val	0	77	165	21	(60)	Tyr[b]	0	14	30
11	(50)	Val	77	165	151						

TABLE 2.6C: - Identification: gas chromatography, and amino acid analysis after acid hydrolysis (underlined yields)

Peptide CNII (SPITC; GLY) 3.8 mg protein program

Residue No.		Residue	nano moles R-1	nano moles R	nano moles R+1	Residue No.		Residue	nano moles R-1	nano moles R	nano moles R+1
1	(86)	Asn	-	109	3	19	(104)	Thr	0	74	13
2	(87)	Ser	0	190	0	20	(105)	Ser	0	37	0
3	(88)	Phe	0	206	13	21	(106)	Ala	4	54	15
4	(89)	Val	0	192	14	22	(107)	Asn	0	42	7
5	(90)	Asn	0	123	14	23	(108)	Arg	13	23	60
6	(91)	Asp[a]	36	121	4	24	(109)	Arg	23	60	28
7	(92)	Val	4	138	17	25	(110)	Ser	0	12	0
8	(93)	Phe	0	197	2	26	(111)	Thr	0	39	11
9	(94)	Glu[a]	0	130	10	27	(112)	Val	2	26	7
10	(95)	Arg	15	82	-	28	(113)	Ser	0	12	16
11	(96)	Ile	4	170	25	29	(114)	Ser	12	16	0
12	(97)	Ala	0	129	17	30	(115)	Arg	11	23	13
13	(98)	Ser	0	92	0	31	(116)	Glu[a]	9	14	-
14	(99)	Glu[a]	0	100	0	32	(117)	Ile[b]	5	13	8
15	(100)	Ala	4	100	15	33	(118)	Gln[b]	0	9	0
16	(101)	Ser	0	111	0	34	(119)	Thr[b]	0	8	6
17	(102)	Arg	11	63	-	35	(120)	Ala[b]	3	5	3
18	(103)	Leu	5	91	28	36	(121)	Val[b]	1	5	2

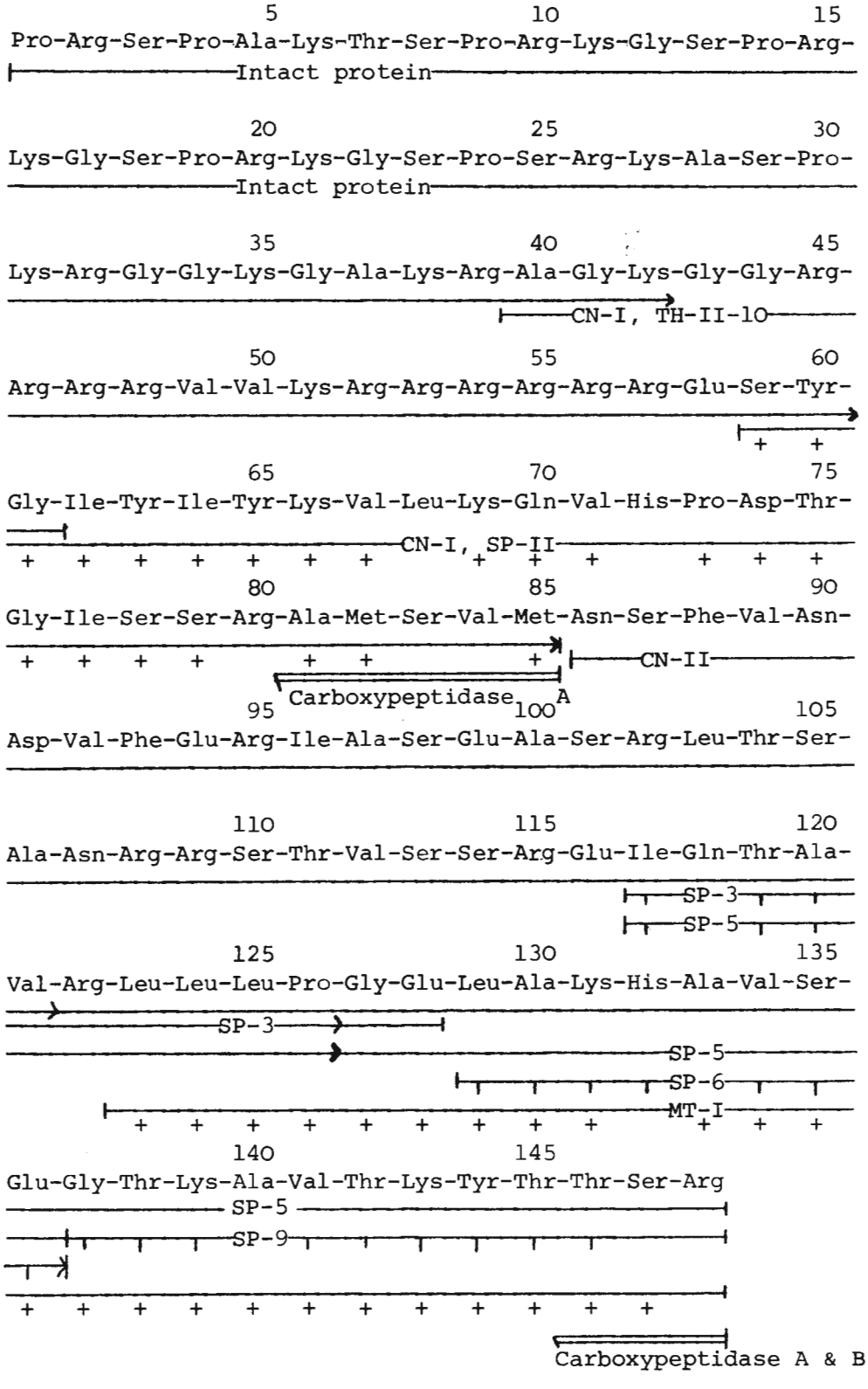


Fig. 2.15: Primary structure of sperm histone H2B (3) Parechinus angulosus

Alignment of the peptides is shown.

› = indicates final residue sequenced in a particular peptide.

—— = liquid-state sequencing

—— = solid-state sequencing
+ +

—— = DANSYL-Edman sequencing

Types of cleavage

CN = Cyanogen Bromide SP = Staphylococcus aureus

TH = Thermolysin MT = Maleylated Trypsin

Quantitative data are available only for the material which was sequenced in the automatic spinning cup sequencer. These sequences comprising 96 residues out of 148 include all the regions of the protein which differ from $H2B_{(2)}$ Parechinus. This is apparent when the amino acid compositions of the peptides are compared to those of $H2B_{(2)}$ (Tables 2.1 and 2.5). The complete sequence of $H2B_{(2)}$ is documented by full quantitation so the use of qualitative sequencing in regions of $H2B_{(3)}$ which are homologous to $H2B_{(2)}$ seems justified.

2.5 PARTIAL SEQUENCE OF $H2B_{(1)}$ PSAMMECHINUS

2.5.1 Generation of CNBr peptides

Due to limited amounts of this protein CNBr cleavage was performed in the cup of the sequencer after prior sequencing of the amino-terminus (4.3.1.1). The amino acid composition of the whole protein is given in Table 2.8.

2.5.2 Partial sequence

The uncleaved protein $H2B_{(1)}$ Psammechinus was sequenced for 50 steps yielding a single sequence (Table 2.7A) and then subjected to CNBr cleavage in the cup. A further 11 amino acid residues following a methionine residue were removed and identified (Table 2.7B). The partial sequence of $H2B_{(1)}$ Psammechinus is given in Fig. 2.16.

2.6 PARTIAL SEQUENCE OF $H2B_{(2)}$ PSAMMECHINUS

2.6.1 Generation and purification of CNBr peptides

Cyanogen bromide digestion was done as for $H2B_{(2)}$ Parechinus (2.2.1.1). Figure 2.17 illustrates the separation of the two large peptides on Sephadex G-100. The amino acid composition of the whole protein and of these peptides is given in Table 2.8.

2.6.2 Partial sequence

The amino-terminal CN-I fragment of histone $H2B_{(2)}$ *Psammechinus* and the carboxy-terminal CN-II fragment were sequentially degraded and 73 residues were assigned. (Table 2.7 D and E, Fig.2.16).

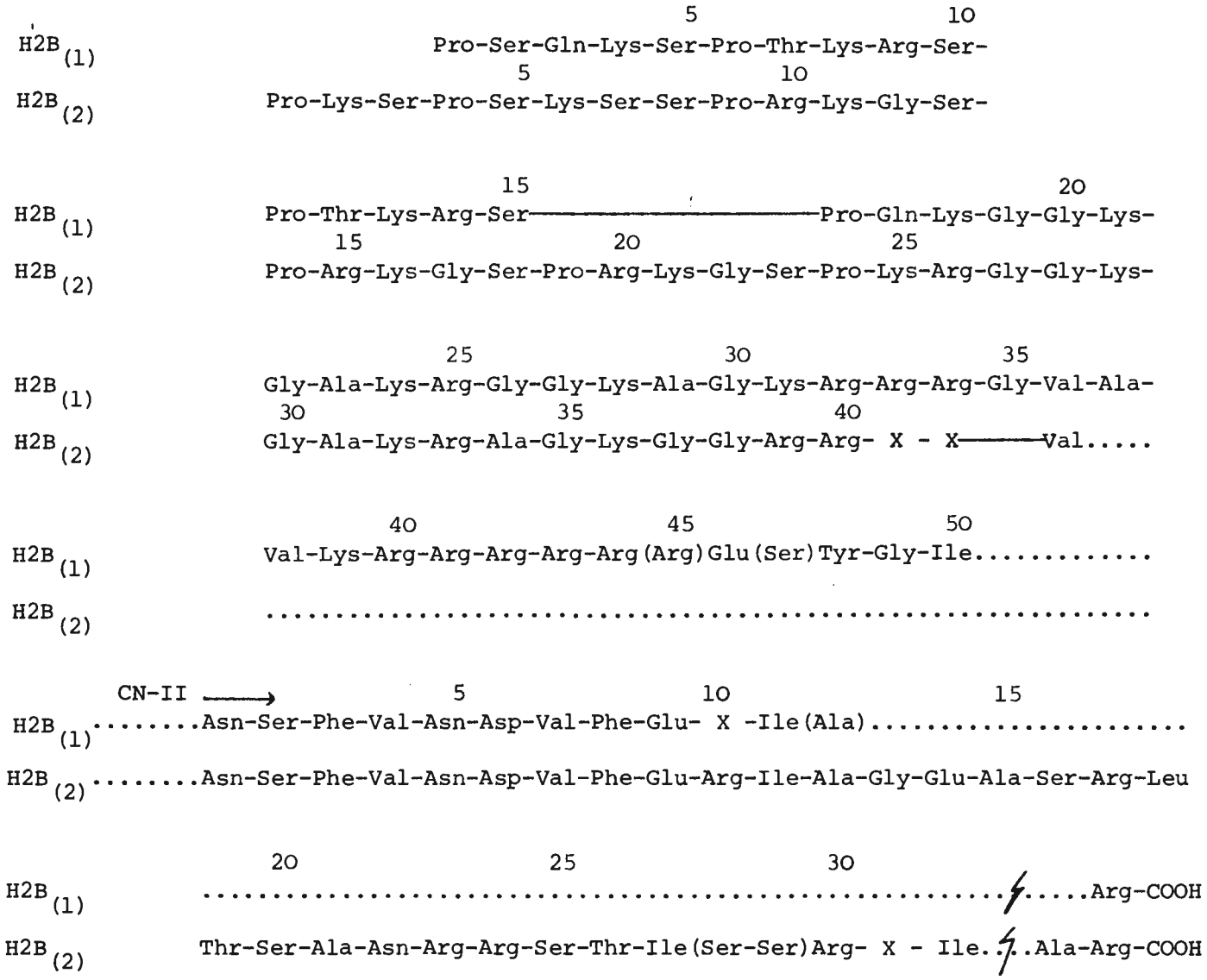


Fig.2.16: Partial primary structure of sperm histones $H2B_{(1)}$ and $H2B_{(2)}$ of Psammechinus miliaris.

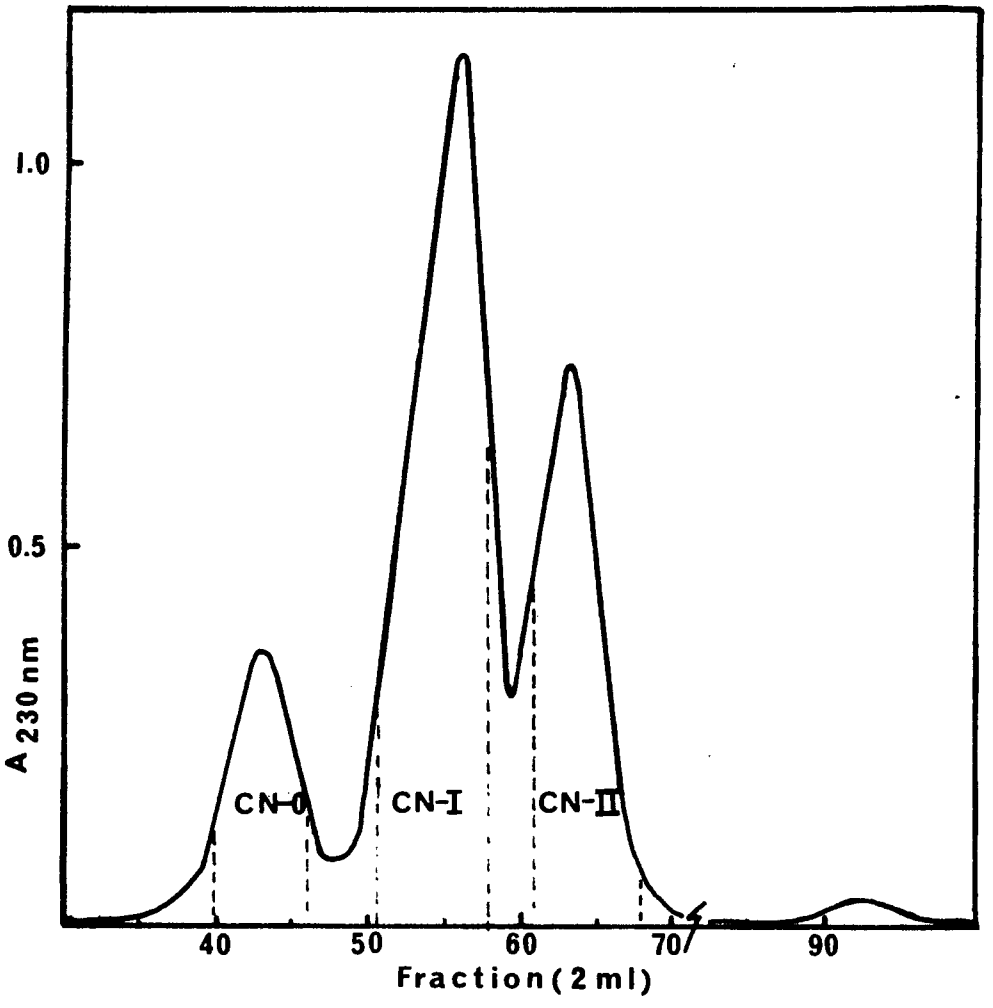


Fig. 2.17: Sephadex G-100 elution pattern of a CNBr digest of 13 mg of $H2B_{(2)}$ Psammechinus. The dried mixture of peptides was dissolved in 1 ml 8 M urea in 10 mM HCl. Eluant 10 mM HCl. Column 16 x 100 mm. Cuvette 1 cm.

TABLE 2.7 Automatic sequencing of histones $H2B_{(1)}$ and $H2B_{(2)}$ of Psammechinus miliaris.

Identification: gas chromatography and amino acid analysis after acid hydrolysis (underlined values). If the amino acid assignment is in brackets, there is only qualitative evidence for that assignment. If an X is present at any position, no assignment could be made.

(SPITC) = 4-sulfophenylisothiocyanate modified material.
(GLY) = aspartyl and glutamyl glycine methylester of the peptide.
R = yield of amino acid derivative assigned to that position.
R-1 = yield of that amino acid in the previous cycle.
R+1 = yield of that amino acid in the following cycle.
( ) = sequence alignment number.
a = proline and theonine are incompletely resolved on G.C. and consequently values given are estimates from peak height. On hydrolysis and amino acid analysis cycles identified as threonine gave a peak near α-amino butyric acid.
b = accompanied by an increase in glycine after acid hydrolysis and amino acid analysis.

TABLE 2.7 A $H2B_{(1)}$ Psammechinus

Intact protein 5.7mg.

Residue No.	Residue	nano moles			Residue No.	Residue	nano moles		
		R-1	R	R+1			R-1	R	R+1
1	Pro				26	Gly	0	33	37
2	Ser	0	105	0	27	Gly	33	37	8
3	Thr		114	-	28	Lys	4	22	-
4	Lys	5	98	-	29	Ala	9	40	22
5	Ser	0	157	0	30	Gly	8	39	11
6	Pro	0	63	60	31	Lys	10	20	22
7	Thr[a]	0	55	10	32	Arg	3	15	29
8	Lys	5	52	11	33	Arg	15	29	32
9	Arg	0	39	-	34	Arg	29	32	11
10	Ser	0	43	8	35	Gly	8	38	14
11	Pro	4	54	51	36	Val	14	43	27
12	Thr[a]	0	50	19	37	Ala	7	38	30
13	Lys	6	42	11	38	Val	27	54	28
14	Arg	0	42	-	39	Lys	6	16	9
15	Ser	0	74	12	40	Arg	11	16	26
16	Pro	4	45	10	41	Arg	16	26	28
17	Gln	0	47	8	42	Arg	26	28	31
18	Lys	12	27	-	43	Arg	28	31	35
19	Gly	6	59	52	44	Arg	31	35	29
20	Gly	59	52	4	45	Arg	35	29	26
21	Lys	9	35	-	46	Glu	2	11	8
22	Gly	4	45	6	47	(Ser)			
23	Ala	0	65	24	48	Tyr	2	9	-
24	Lys	9	27	12	49	Gly	1	10	-
25	Arg	0	30	-	50	Ile	1	12	-

TABLE 2.7 B $H2B_{(1)}$ Psammechinus

Blocked protein after CNBr cleavage in sequencer cup (see text).

Residue No.	Residue	nano moles R-1	nano moles R	nano moles R+1	Residue No.	Residue	nano moles R-1	nano moles R	nano moles R+1
1 (91)	Asn	-	30	2	7 (97)	Val	0	12	4
2 (92)	Ser	0	22	0	8 (98)	Phe	0	8	0
3 (93)	Phe	0	16	0	9 (99)	Glu	0	5	0
4 (94)	Val	0	16	0	10 (100)	X		-	
5 (95)	Asn	0	9	5	11 (101)	Ile	0	6	0
6 (96)	Asp	2	10	4	12 (102)	(Ala)			

TABLE 2.7 C $H2B_{(2)}$ Psammechinus

Peptide CN-I 2.8mg

Residue No.	Residue	nano moles R-1	nano moles R	nano moles R+1	Residue No.	Residue	nano moles R-1	nano moles R	nano moles R+1
1	Pro	-	110	11	23	Ser	0	39	0
2	Lys	0 10	51 64	0	24	Pro	5	17 10	7
3	Ser	0	128	0	25	Lys	7	21	12
4	Pro	0	148	8	26	Arg	9	22	7
5	Ser	0	185	0	27	Gly	0	23	23
6	Lys	0 5	77 36	0	28	Gly	23	23	17
7	Ser	0	136	153	29	Lys	7	14	6
8	Ser	136	153	0	30	Gly	17	36	22
9	Pro	0	42 26	13	31	Ala	3	30	9
10	Arg	6	43	10	32	Lys	0 6	7 11	0 8
11	Lys	0 9	33 25	0	33	Arg	2	12	8
12	Gly	0	77	17	34	Ala	8	24	10
13	Ser	0	96	0	35	Gly	5	11	4
14	Pro	5	37 20	13	36	Lys	5	9	5
15	Arg	2	29	11	37	Gly	4	13	14
16	Lys	0 6	12 25	4	38	Gly	13	14	9
17	Gly	0	70	10	39	Arg	6	15	18
18	Ser	0	37	0	40	Arg	15	18	13
19	Pro	6	28 13	10	41	(Arg)	18	13	11
20	Arg	2	15	10	42	X			
21	Lys	0 11	10 21	9 0	43	Val	2	7	-
22	Gly	0	20	8					

TABLE 2.7 D $H2B_{(2)}$ Psammechinus

Peptide CN-II (SPITC; GLY) 2.4mg

Residue No.	Residue	nano moles			Residue No.	Residue	nano moles		
		R-1	R	R+1			R-1	R	R+1
1 (91)	Asn	-	100		17 (107)	Arg	10	47	-
2 (92)	Ser	0	95	13	18 (108)	Leu	4	67	20
3 (93)	Phe	0	156	9	19 (109)	Thr	0	55	18
4 (94)	Val	0	153	10	20 (110)	Ser	5	38	7
5 (95)	Asn	0	147	39	21 (111)	Ala	7	38	12
6 (96)	Asp	40	22 93[b]	7	22 (112)	Asn	0	19	3
7 (97)	Val	0	126	15	23 (113)	Arg	8	21	21
8 (98)	Phe	4	133	14	24 (114)	Arg	21	21	12
9 (99)	Glu	0	14 76[b]	0	25 (115)	Ser	5	15	6
10 (100)	Arg	10	60	-	26 (116)	Thr	6	17	8
11 (101)	Ile	0	103	14	27 (117)	Ile	3	20	5
12 (102)	Ala	3	82	15	28 (118)	(Ser)	-	-	-
13 (103)	Gly	0	98	13	29 (119)	(Ser)	-	-	-
14 (104)	Glu	0	49[b]	0	30 (120)	Arg	5	9	6
15 (105)	Ala	6	92	20	31 (121)	(Glu)	-	-	-
16 (106)	Ser	0	89	16	32 (122)	Ile	5	12	5

TABLE 2.8:

	$H2B_{(1)}$ Psammechinus	$H2B_{(2)}$ Psammechinus		
	Whole protein	Whole protein	CN-I	CN-II
Aspartic acid	6.0	6.0	2.0	3.7
Threonine	9.9	7.0	1.0	6.2
Serine	12.9	15.7	8.9	6.0
Glutamic acid	10.2	8.2	2.2	5.9
Proline	5.9	8.3	7.2	1.2
Glycine	12.2	15.1	12.4	3.2
Alanine	10.2	10.4	2.2	7.7
Valine	12.2	9.8	4.0	5.0
Methionine	1.8	1.8	*	0
Isoleucine	4.0	5.9	2.8	2.7
Leucine	6.1	6.2	1.1	4.8
Tyrosine	4.9	3.6	2.9	1.0
Phenylalanine	2.1	2.0	0	1.9
Lysine	14.4	15.8	12.0	3.1
Histidine	2.1	1.9	1.0	1.0
Arginine	20.2	23.8	14.5	7.9

Amino acid composition (number of residues assuming two phenylalanine residues) of $H2B_{(1)}$ and $H2B_{(2)}$ from sperm of Psammechinus miliaris. The composition of the cyanogen bromide peptides of $H2B_{(2)}$ Psammechinus is also given. No corrections have been made for losses during or incompleteness of hydrolysis.
*Detected as homoserine and homoserine lactone but not quantitated.

PART 3

3.1 BASIC SPERM PROTEINS

The basic proteins associated with DNA in sperm cells vary considerably. To pack the DNA tightly and protect it from the environment a considerable number of different proteins appear to be suitable. In general, three main types of basic sperm proteins have been observed in different organisms. These are small basic proteins (30-80 amino acids), large basic proteins, and histone-like proteins.

Small basic proteins, the protamines, are found in the sperm of many species of fish and have been best characterized in trout, salmon and herring (for a review see Ando et.al., 1973). These proteins are extremely rich in arginine and completely replace somatic type histones as spermatogenesis proceeds. Protamine type proteins have also been observed in sperm of chicken (Nakano et.al., 1970) squid and octopus (Subirana et.al., 1973) and many other organisms (Bloch, 1969).

Small arginine-cysteine rich "protamines" which also replace somatic type histones are found in mature sperm of eutherian mammals. The protein from bull sperm has been sequenced (Coelingh et.al., 1972). The rat sperm protein, apparantly belonging to the same type of protein, and several other basic proteins characteristic of various stages of spermatogenesis in rat have been extensively studied electrophoretically and by amino acid analysis. (Kistler et.al., 1973; Platz et.al., 1977). A histone H2B, containing cysteine, has also been identified from rat testis (Shires et.al., 1976) but there is no evidence that this protein is present in sperm cells.

A large basic protein containing 43 mole % arginine has been found in sperm of the limpet Patella vulgata (Subirana et.al., 1973). A similar protein has been observed in sperm of a local species of limpet Patella granatina (W.F. Brandt, unpublished). Such proteins have been reported for many organisms on the basis of electrophoretic studies (Hnilica, L.S., 1972). These large arginine rich proteins seem to vary considerably in composition and appear in general to replace somatic type histones in the mature sperm cell in a variety of marine invertebrates. Histone-like proteins have been identified by electrophoretic criteria in the sperm cells of many organisms. Included in this class are echinoderms; fish such as carp, barbel and toadfish; and certain frogs (Subirana and Palau, 1968; Subirana et.al., 1975). Subirana (1975) describes various differences in the electrophoretic pattern of these proteins from the organisms investigated. In particular, the histone complement from sperm of sea urchin is described as having additional components "like histone H1 and histone H2B". We have now identified multiple H2B histones in sperm of two species of sea urchin.

In Psammechinus miliaris two sperm H2B histones have identical electrophoretic mobilities. In Parechinus angulosus three H2B variants have been identified; one of these, $H2B_{(2)}$, co-electrophoreses with H3 while the other, two $H2B_{(1)}$ and $H2B_{(3)}$, migrate together slightly slower than H3. (Fig. 1.4, Fig. 1.5). To complicate the matter, similar H2B-like proteins with electrophoretic mobilities even slower than the $H2B_{(1)}$ - $H2B_{(3)}$ pair but faster than sperm H1 have been noted in testes of P. angulosus (Fig.1.1, Table 1.2) and in gonads of P. miliaris (Wouters-Tyrou, 1977). This clearly establishes the unreliability of identification of histones solely by electrophoretic criteria.

The complete sequences are available for H2B histones from calf thymus (Iwai et.al., 1972), trout testis (Kootstra and Bailey, 1976), and limpet Patella granatina (van Helden,et.al.,1978b). Partial sequences for H2B histones from nucleated erythrocytes of various animals have been reported recently (van Helden et.al.,1978a). If the C-terminal regions of the sperm H2B histones which I have investigated are compared to the somatic histones H2B, close homology becomes apparent. However, the N-terminal regions are variable. In general, the proteins from calf, trout, and the nucleated erythrocytes from different taxonomic vertebrate classes have similar N-terminal sequences, while the N-terminal sequences of the H2B histone from the mollusc P. granatina and those from the sea urchin H2B histones reported in the present study are different from each other and from that of the vertebrates. Partial sequences of H2B histones from sea urchin embryos (Brandt, et.al., 1978) indicates the presence of multiple forms of H2B even in a single species. The N-terminal regions of embryo H2B histones are different from the sperm H2B histones of the same organism and also from the histones H2B from fully differentiated diploid vertebrate cells,but again homology is observed in the C-terminal parts. These results substantiate claims based on electrophoretic data on the existence of variants of H2B in sea urchin embryos (Cohen et al., 1975). Though more sequence data is desirable, there can be no doubt that H2B histones are considerably more variable than previously supposed. In particular, H2B histones may vary in different types of tissue in the same organism and it is therefore very important in structural studies that the source of the histone is carefully controlled and specified.

Within the context of variability of the general types of basic sperm proteins the even more variable H2B histones of sea urchin will be considered and I hope to show that evolution of basic sperm proteins has proceeded at a much more rapid rate than that observed for somatic histones which are considered to have evolved at a rate of less than 1 accepted point mutation per 100 million years (Dayhoff, 1976).

3.2 COMPARISON OF STRUCTURE AND EVOLUTION OF SEA URCHIN SPERM H2B HISTONES

The histones H2B from sperm cells of the two sea urchin species Parechinus angulosus and Psammechinus miliaris are closely related proteins. The amino-terminal regions of these proteins are elongated by various duplications, compared to histone $H2B_{calf}$ (Iwai et.al., 1972) (Fig.3.1). A pentapeptide repeat is the most prominent of these duplications though there is also evidence of a repeating tripeptide (alignment positions 33-48, Fig. 3.1). In two of the sperm histones ($H2B_{(1)}$ Psammechinus and $H2B_{(1)}$ Parechinus) the amino-terminal sequences begin with a pentapeptide. In the remaining three histones ($H2B_{(2)}$ Psammechinus, $H2B_{(2)}$ Parechinus and $H2B_{(3)}$ Parechinus) the pentapeptide region is preceded at the amino-terminal side by a tripeptide. The first pentapeptide in all sperm histones differs from the subsequent repeats. The reiterated pentapeptide is either Pro-Thr-Lys-Arg-Ser ($H2B_{(1)}$ Psammechinus, $H2B_{(1)}$ Parechinus) or Pro-Arg-Lys-Gly-Ser ($H2B_{(2)}$ Psammechinus, $H2B_{(2)}$ Parechinus, $H2B_{(3)}$ Parechinus). The reiteration occurs twice in $H2B_{(1)}$ Psammechinus, $H2B_{(2)}$ Parechinus and three times in $H2B_{(1)}$ Parechinus, $H2B_{(2)}$ Psammechinus, $H2B_{(3)}$ Parechinus. A reiterated pentapeptide structure is also present in the amino terminal sequence of an H2B from sperm of the irregular sea urchin Echinolampas crassa (Fig.3.1, W.N. Strickland, unpublished). Since the identical reiterated pentapeptide units are the most constant features in the N-terminal parts of the sperm H2B histones, one could postulate an ancestral pentapeptide with divergence to these two pentapeptides, characteristic for either the $H2B_{(1)}$ or the $H2B_{(2)}$ type.

A scheme for evolution of the pentapeptide region is presented in Fig. 3.2. Single base substitutions, deletions, duplications and insertions are each

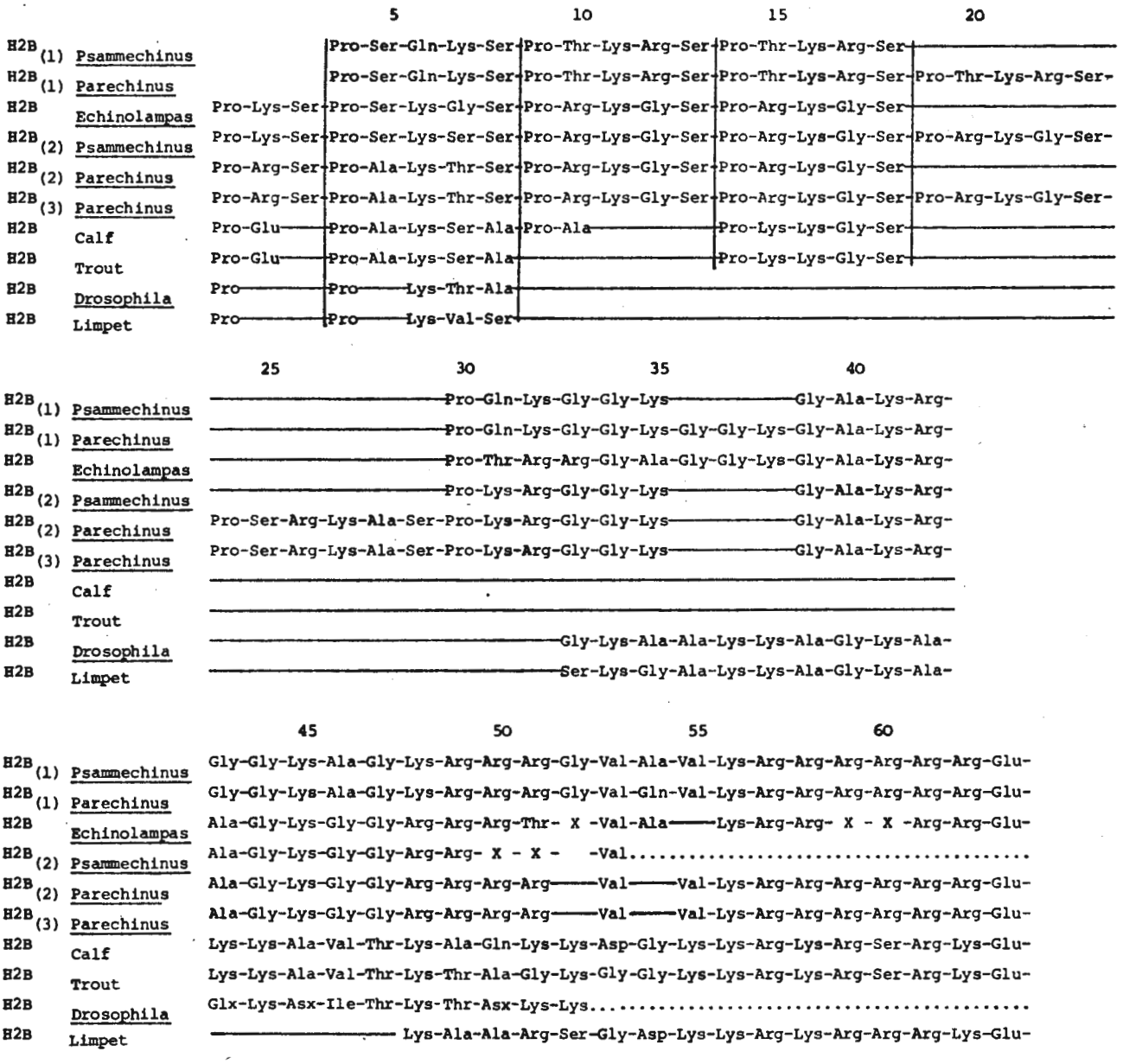


Fig. 3.1: Amino acid sequences of the variable amino terminals of various H2B histones from Parechinus angulosus, Psammechinus miliaris (part 2 of this thesis), Echinolampas crassa (W.N. Strickland unpublished), calf (Iwai et.al., 1972), Trout (Kootstra and Bailey, 1976), Drosophila (Elgin and Weintraub, 1975), and Mollusc (van Helden, 1978).

——— = gaps for alignment | | = pentapeptides ..... not sequenced.

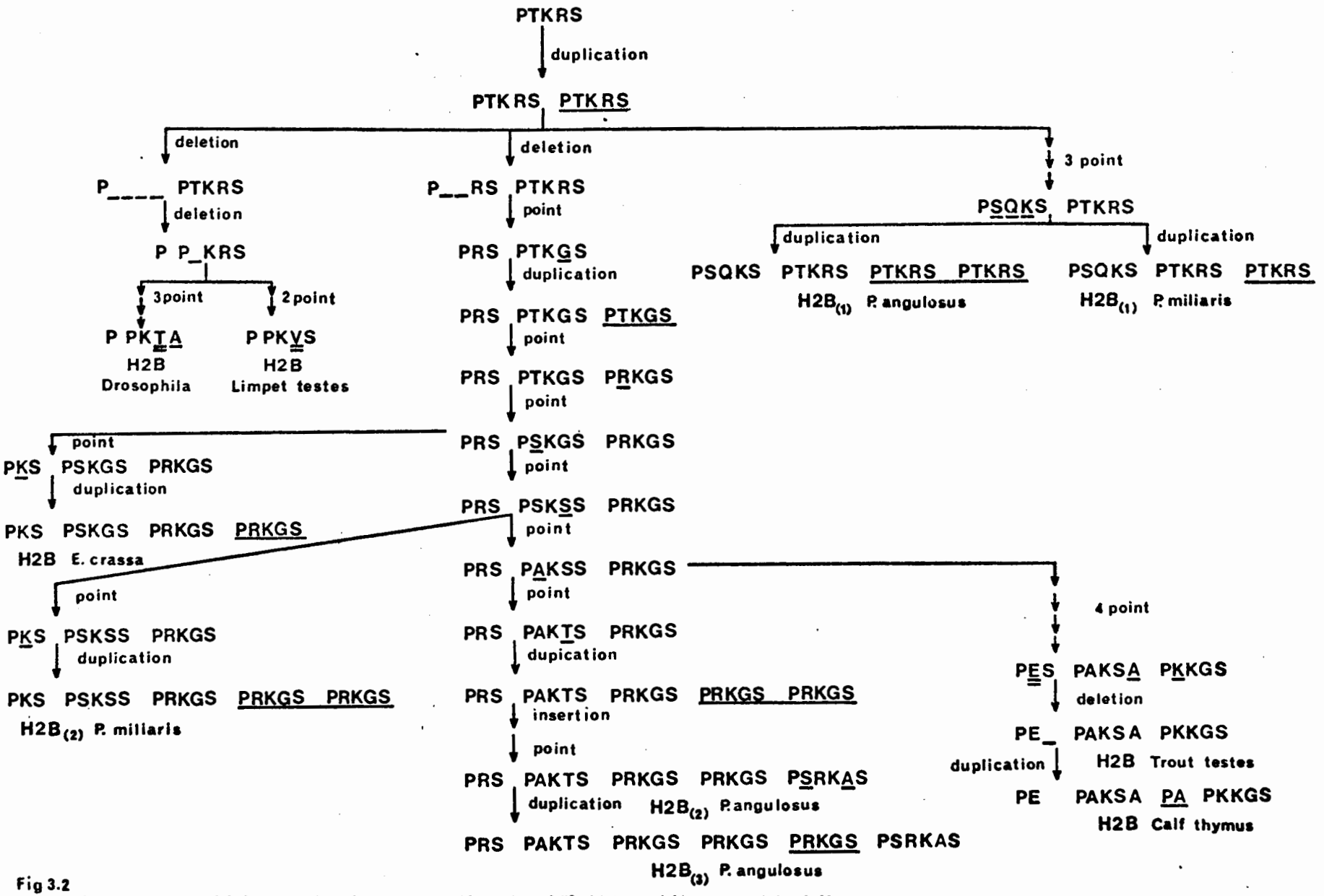


Fig 3.2

A single letter notation model for evolution of the pentapeptide region of H2B histones (alignment position 1-29, Fig.3.1). The pentapeptide seen from alignment position 9-13 in the $H2B_{(1)}$ type proteins (Fig. 3.1) is considered as the ancestral pentapeptide. Considering the known structures for $H2B_{(1)}$ histones an evolution scheme by duplication, deletion, and the least number of point mutations is proposed. The assumption is made that point mutations are equally likely at all positions in a codon. Arrows indicate mutational events and the changes are underlined in the subsequent sequence. The branching points and length of branches are intended to show general relationships but no time scale of divergence is implied. The codon for arginine in the ancestral pentapeptide is assumed to be CGU or CGC.

considered as single mutational events. The branch points are postulated on the basis that divergence from common sequences occurred as the result of the least number of mutational events. If one examines the amino acids present at positions 1-5 in the pentapeptide, positions 2 and 4 are found to be quite variable while positions 1, 3 and 5 are more constant. Proline is the only amino acid seen at position 1. Position 3 usually contains lysine and only occasionally glutamine. Position 5 is usually serine but alanine may be present. Threonine at position 2 would allow mutation to all other amino acids seen at that position by single point mutations. The same is true for Arg in position 4 with the exception of threonine in that position in Drosophila. From these observations the pentapeptide Pro-Thr-Lys-Arg-Ser which is the reiterated pentapeptide in $H2B_{(1)}$ type proteins (Fig. 3.1) is proposed as the ancestral pentapeptide. Duplication of the ancestral pentapeptide with subsequent point mutations leads to the precursor of histone $H2B_{(1)}$ whereas partial deletion could have produced the precursor protein for the $H2B_{(2)}$ type of protein (Fig. 3.2). In each evolutionary branch, point mutations and further duplications then yielded the present H2B proteins of sea urchin sperm. For example, $H2B_{(3)}$ Parechinus differs from histone $H2B_{(2)}$ Parechinus by the presence of an additional pentapeptide and by a single substitution at alignment position 103 (Fig. 3.3). The pentapeptide, Pro-Arg-Lys-Gly-Ser, is repeated three times in histone $H2B_{(3)}$ Parechinus while in histone $H2B_{(2)}$ Parechinus there are only two repeats of that particular pentapeptide. In both proteins this region is followed by a hexapeptide also beginning with proline and ending with serine. Mutation and insertion of serine within one of the reiterated pentapeptides could have yielded this hexapeptide seen between positions 24-29. Because $H2B_{(3)}$ Parechinus has three identical pentapeptide repeats and $H2B_{(2)}$ only two identical repeats, and both contain the same hexapeptide; it seems likely that the previously postulated mutation of one pentapeptide unit to a hexapeptide occurred prior to the

divergence of the ancestral polypeptide into the two proteins. Deletion or insertion of a repeating unit could have been the mechanism for this divergence.

In some cases more than one branching point or common intermediate is possible. For instance, $H2B_{(2)}$ of P. miliaris and the E. crassa H2B share a common N-terminal sequence Pro-Lys-Ser while $H2B_{(2)}$ and $H2B_{(3)}$ of P.angulosus have the sequence Pro-Arg-Ser. In the sequence of the 1st complete pentapeptide, however, P.miliaris and P.angulosus H2B proteins could be regarded as having a common intermediate sequence (Pro-Ser-Lys-Ser-Ser) with the Pro-Lys-Ser sequence in P.miliaris evolving independantly of that sequence in E.crassa. Another possible branching arrangement would have E.crassa and P.miliaris structures diverging from the main branch together with subsequent modification of the 1st complete pentapeptide Pro-Ser-Lys-Gly-Ser of E. crassa to the Pro-Ser-Lys-Ser-Ser of P.miliaris by mutation of Gly to Ser. Each branching order involves the assumption of two mutational events. The construction of the branching in Fig. 3.2 was made on the basis that P. miliaris or P. angulosus are more closely related to each other than to E. crassa. (Hyman, 1955; Clark and Courtman-Stock, 1976).

Remnants of a pentapeptide repeat can also be seen in the somatic H2B histones from bovine thymus (Iwai, 1972) and the erythocytes of Gallus domesticus, Xenopus laevis and Crocodilus niloticus (van Helden et.al., 1978a) namely residues 3 to 7 and 10 to 14. A similar pentapeptide repeat is also found in histone H2B from trout testes from residue 3 to 7 and 8 to 12 (Kootstra and Bailey, 1976). H2B histones from Drosophila (Elgin and Weintraub, 1975) and the limpet Patella granatina (van Helden et.al.,1978b) have amino terminal sequences which could be regarded as remnants of a pentapeptide. The amino terminal regions of H2B histones from sea urchins,

calf, trout, Drosophila and limpet are aligned for homology in Fig. 3.1 and possible evolutionary relationships are indicated in Fig. 3.2. The amino terminal sequences of H2B from chicken, toad, and crocodile are similar to calf H2B. The observation that the amino terminal sequences of H2B from a mollusc and an insect are more similar to each other than to homologous sequences found in echinoderms and vertebrates is in agreement with a generally accepted evolutionary tree. (Stafford, 1960).

Protamines and amino terminal regions of the sperm H2B histones show certain structural parallels. Both types of protein contain proline and hydroxamino acids separated by basic amino acids arranged as a pentapeptide as outlined elsewhere (von Holt, et al., 1978). The sperm H2B histones also contain an arginine rich block (alignment position 57-62, Fig. 3.1) which may be homologous to several poly arginine sequences in protamines. (Dayhoff, 1972). These structural parallels may indicate an evolutionary link between protamines and sea urchin sperm histones; but also in view of the apparently rapid evolution of sperm proteins, may only reflect convergent evolution.

Rapid evolution of the sperm proteins especially in regard to the number of repeats may well be the result of adjustments to environmental conditions in order to maintain an optimal chromatin structure. In P.angulosus the amounts of $H2B_{(2)}$ and $H2B_{(3)}$ varies, (Fig. 1.5), depending on the habitat of the animals. In the population existing on the eastern shores of the Cape Peninsula, no $H2B_{(3)}$ has been detected; while in another population located at 12 meters depth on the western side, the incidence is 95% $H2B_{(3)}$ to 5% $H2B_{(2)}$. In two other locations on the western side of the Cape Peninsula (both at the intertidal zone) the ratio of $H2B_{(2)}$ to $H2B_{(3)}$ is about 50 : 50.

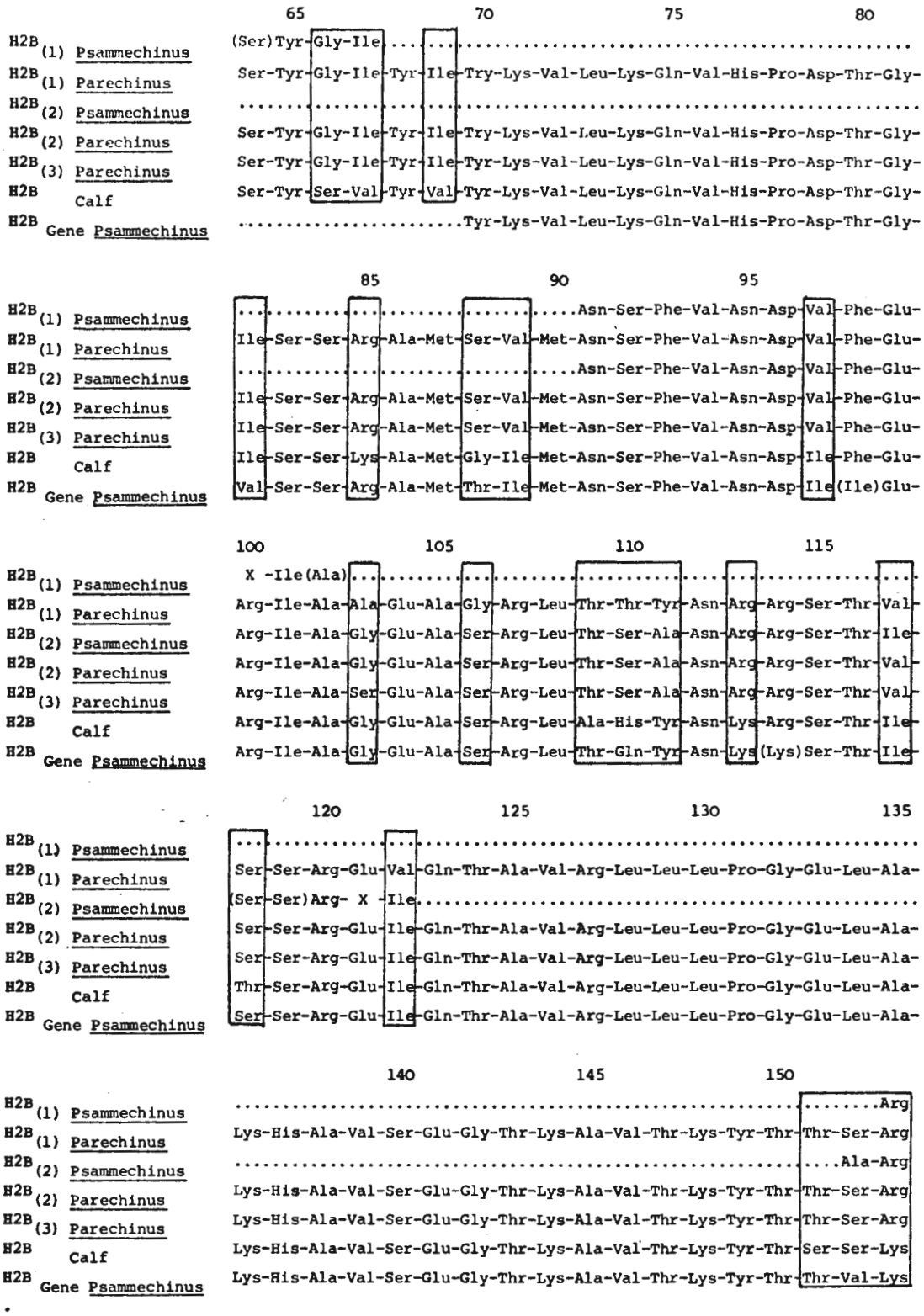


Fig. 3.3: Amino acid sequences of the 90 C-terminal amino acids of histones $H2B_{(1)}$ and $H2B_{(2)}$ from sperm of Psammechinus miliaris and those of the three sperm histones H2B of Parechinus angulosus. These sequences are compared to that of calf thymus H2B (Iwai et.al., 1972). Also indicated is the sequence determined from DNA from Psammechinus miliaris (Birnstiel et.al., 1977) beginning at alignment position 70. ..... not sequenced. ☐ = variable positions.

While the amino terminal portions of the sea urchin sperm histones H2B have undergone rapid evolution, the region from alignment position 63 onwards has been conserved. This region has remained virtually unchanged when calf thymus H2B is compared to the sea urchin proteins (Fig. 3.1 and 3.3). Certain changes at various positions in this otherwise conservative part of the protein do, however, occur. For instance, there are several isoleucine ⟷ valine interchanges. At alignment position 103 a different amino acid (alanine, glycine or serine) has been found in each of the three Parechinus H2B histones. It is interesting that serine (a variant of the usual glycine) has been observed in the corresponding position in an H2B from mouse ascites tumor (Franklin and Zweidler, 1977). The more conservative part of the sperm H2B histones do vary only slightly from each other and from other H2B histones. This conservation may reflect complex protein-protein interactions in the histone DNA complex which must be preserved while the variation allows for subtle conformational changes necessary for the optimal conformation of the complex of any particular cell type.

When the sequence of the carboxyl-terminal 84 amino acids, deduced from the nucleotide sequence of the gene for an H2B histone from Psammechinus miliaris as determined by Birnstiel et.al., (1977) (i.e., alignment positions 70-153 Fig. 3.3) is compared with the known sequence of all sea urchin sperm histones H2B, identity is observed in 71 positions. In nine of the thirteen variable positions the amino acid given by the gene is present in at least one of the sea urchin histone variants. There are 8 non-identical positions between the sequence of the somatic histone H2B from calf and the amino acid sequence of the sea urchin gene in the same region.

In five positions of sea urchin H2B histones interchanges of isoleucine and valine have occurred (alignment positions 82, 89, 97, 117, 122). Other interchanges involve lysine to arginine (113, 153), threonine to serine (88), glycine to serine or alanine (103), serine to glycine (106), glutamine to serine or threonine (110), tyrosine to alanine (111), and valine to serine or alanine (152). The variations seen from alignment position 63 onwards are clustered in certain regions. Most of the variations occur from alignment position 85-122 and in that region variation occurs more often at some positions than others. The remaining 31 positions are invariant except for the last three residues. When the carboxy-terminal sequences for other H2B histones are compared to those given in Fig. 3.3 the same variable positions are noted.

To appreciate the speed of evolutionary change in the sperm histones H2B the occurance of six variant structures in only three species of the Class Echinoidea comprising about 800 species (Barnes, 1974) must be seen against the virtual constancy of the histone H2B structure in fully differentiated somatic cells of vertebrate classes which have evolved several hundred million years apart from each other (van Helden et.al., 1978a; Dayhoff, 1976).

PART 4

4.1 PURIFICATION OF SPERM H2B HISTONES

4.1.1 Selective extraction of nucleoprotein

Sperm cells were collected by placing sexually mature male gonads in a muslin bag and gently filtering out the sperm by washing with sea water at ambient temperature. After centrifugation (10 000 g for 10 min) they were resuspended in ice cold 0.15 M NaCl and homogenized with the tight plunger of a Dounce tissue homogenizer. The insoluble material was pelleted at 6 000 g for 10 minutes and the pellet resuspended in 0.15 M NaCl. This procedure was repeated six or seven times.

The histones were then selectively extracted from the pellet by a modification of Johns (1964; 1977) procedure. All extraction steps involved homogenization of the precipitate in the Dounce homogenizer with the extraction medium. Double distilled water (2-3 ml per g wet weight of sperm) was added to the washed crude chromatin and the mixture was left to swell for one hour to facilitate subsequent extractions. With continual stirring, an equal volume of 10% perchloric acid (w/v) was added, homogenized and the mixture stirred for 30-45 minutes. Following centrifugation, the precipitate was extracted while stirring for 18-20 hours with a solution of 0.25 N HCl in 80% ethanol (2-3 ml ethanolic HCl per g wet weight of sperm). The precipitate was re-extracted in the same volume of ethanolic HCl for 60 minutes. These steps removed the bulk of the histones except H2B from the chromatin. Finally, with continual stirring, the crude H2B fraction was extracted for 60 minutes with (2-3 volumes) 0.25 N HCl. All of the above manipulations were done at $+4^{\circ}C$.

4.1.2 Column chromatography

The efluent from columns was monitored at 206 nm on a LKB UVICHORD III absorptiometer or by measuring individual fractions at 230 nm. Selected fractions were pooled, dialyzed and freeze dried.

4.1.2.1 Gel exclusion chromatography on Biogel P60

To remove contaminating histones H1, H2A and H3 as well as unidentified proteins (slow migrating on polyacrylamide gel electrophoresis- Fig.1.2, gel 5), approximately 300 mg crude H2B were placed on a P-60 Biogel column (50 mm x 900 mm) and eluted with 0.01 N HCl, 0.05 M NaCl (Böhm et. al., 1973) at room temperature (Fig.1.3A).

4.1.2.2 Ion exchange chromatography

The H2B fraction from P60 containing a mixture of H2B histones was fractionated on a carboxymethyl cellulose (Whatman CM-52) column (26 mm x 400 mm) with a linear NaCl gradient. The gradient buffers contained 50 mM Na-acetate/HCl pH 4.5 and 6 M urea. A typical elution pattern is shown in Fig.1.3B. The flow rate was maintained using a peristaltic pump. After every run the column was washed with 0.5 M NaCl in urea buffer. All urea solutions were prepared immediately before use to reduce cyanate ion build up.

4.2 CHARACTERIZATION OF H2B HISTONES

4.2.1 Polyacrylamide gel electrophoresis

The electrophoretic method is based on that of Panyim and Chalkley (1969). Gels were run at 2 mA per tube for 3½ hours with 0.9 M acetic acid in the electrode chambers. The gels contained 15 % acrylamide, 0.1 % bis-acrylamide, and 2.5 M urea.

The following solutions were made up:

A. 60% (w/v) acrylamide, 0.4% (w/v) N,N'-methylenebisacrylamide

B. 4% (v/v) N,N,N',N'-tetramethylethylenediamine (TEMED) in 43.2% (v/v) glacial acetic acid

C. 0.2% (w/v) ammonium persulphate in 4 M urea

D. 0.9 M glacial acetic acid (tray buffer)

All solutions were stored at 4°C except solution C which was prepared immediately before use.

Gels were prepared by mixing solutions A, B and C in the ratio 2:1:5. The mixture was degassed under vacuum for 1 minute and poured quickly (to about 90 mm) into 100 x 5 mm glass tubes, which were sealed at the lower end with parafilm. About 200 µl of distilled water was carefully layered onto the polyacrylamide. Polymerization was allowed to occur for 1-2 hours at room temperature. Gels were pre-electrophoresed at 2 mA per gel until the voltage stabilized (usually 3-4 hours).

Histone samples were routinely dissolved in 1 mg/ml in 8 M urea. To reduce disulfide bridges the sample was dissolved 10 mg/ml in 50 mM Tris-HCl pH 8 + 10 mM DTE and allowed to stand at room temperature for 15 minutes; the protein sample was then diluted to 1 mg/ml with 8 M urea. Samples containing 5-40 µg of proteins were applied to each gel and electrophoresis was performed at 2 mA per gel for 3½ hours. About 5 µg of material in any one electrophoretic band usually gave good results. Gels were stained for 1 hour in 0.1% amido black in 0.9 M acetic acid. Destaining was performed in a transverse destainer using 0.9 M acetic acid. Gels were stored in 0.9 M acetic acid in 12 x 150 mm test tubes sealed with parafilm.

Gels were placed on an opaque light box and photographed with a Pentax camera fitted with a 50mm macro lens and a vivitar red filter (25A). Ilford FP4 black and white film was used at 1/15 second at F8. The exposure must be varied depending on the light box used. Film was developed with Acutol diluted 1 + 10 for 6 minutes at 20°C. Negatives were printed on high contrast paper.

4.2.2 End group analysis

4.2.2.1 Dansylation of proteins and peptides

The labelling methods are modified from that of Gray (1972). The test tubes used were cleaned and heated in a glass blower's oven overnight before use.

4.2.2.1.1 Protein labelling

1. Dissolve 50 - 250 µg protein in a 6 x 50 mm test tube in 50 µl of 1% SDS by heating in a boiling water bath for 2-5 min. Cool.
2. Add 50 µl N-ethylmorpholine - mix (N-ethylmorpholine should be colourless; if yellow, re-distil)
3. Add 50 µl DNS-chloride dissolved in 50 mg /ml immediately before use in water free acetone - mix.
4. Cover and incubate 30 minutes at 45°C.
5. Fill tube with acetone to precipitate protein-centrifuge.
6. Carefully decant supernatant (discard) and wash pellet with 80% acetone-centrifuge.
7. Discard supernatant and dry pellet under vacuum.
8. Add 20 µl 5.7 N HCl, seal tubes in a gas flame and incubate at 105°C for 6 - 15 hours. The shorter time gives a better recovery of proline but less complete protein hydrolysis. A 15 hour hydrolysis time was routine.

4.2.2.1.2 Peptide labelling

1. 0.5-5 n moles peptide in a 4 x 50 mm test tube dissolved in 5 µl 0.2 M $NaHCO_3$ (pH must be 8.5-9.)
2. Add 5 µl DNS-chloride dissolved 2.5 mg/ml in acetone (can be kept several weeks as long as not exposed to water, then white DNS-OH forms which is insoluable in acetone).
3. Cover and incubate 30 minutes at 45°C.
4. Dry entire incubate under vacuum.
5. Hydrolyze in 10 µl 5.7 N HCl as for proteins.

4.2.2.1.3 Identification of DNS-derivatives

Stock solutions of standard amino acids were made 1 mg/ml in 95% ethanol (Pro, Ser, Thr, di DNS-Lys, and Arg) or acetone (Ala, Glu, Gly, Ile). For use, 20 µl of the stock solutions desired were mixed and the total volume made up to 400 µl. One spot of this mixture was spotted on the reverse side of a 5 x 5 cm micropolyamide sheet (Schleicher and Schüll) and compared to the unknown spots after development using the four solvent systems described by Narita et al., (1975) (Fig.4.1).

Solvent 1 1.5% formic acid in water (v/v)

Solvent 2 Benzene - acetic acid (9:1 v/v)

Solvent 3 Ethyl acetate-acetic acid-methanol (20:1:1 v/v)

Solvent 4 50 mM Na_3PO_4 in 25% aqueous ethanol.

After hydrolysis of the peptide or protein sample, the dried residue is dissolved in 5 µl of 95% ethanol and spotted onto the polyamide sheet. The sheet is examined under an ultraviolet lamp after development in solvents 1 and 2. Resolution of DNS-glutamic acid and DNS-aspartic acid; of DNS-serine and DNS-threonine; and of DNS-alanine and DNS-NH_2 is achieved after development in solvent 3. Development in solvent 4 separates

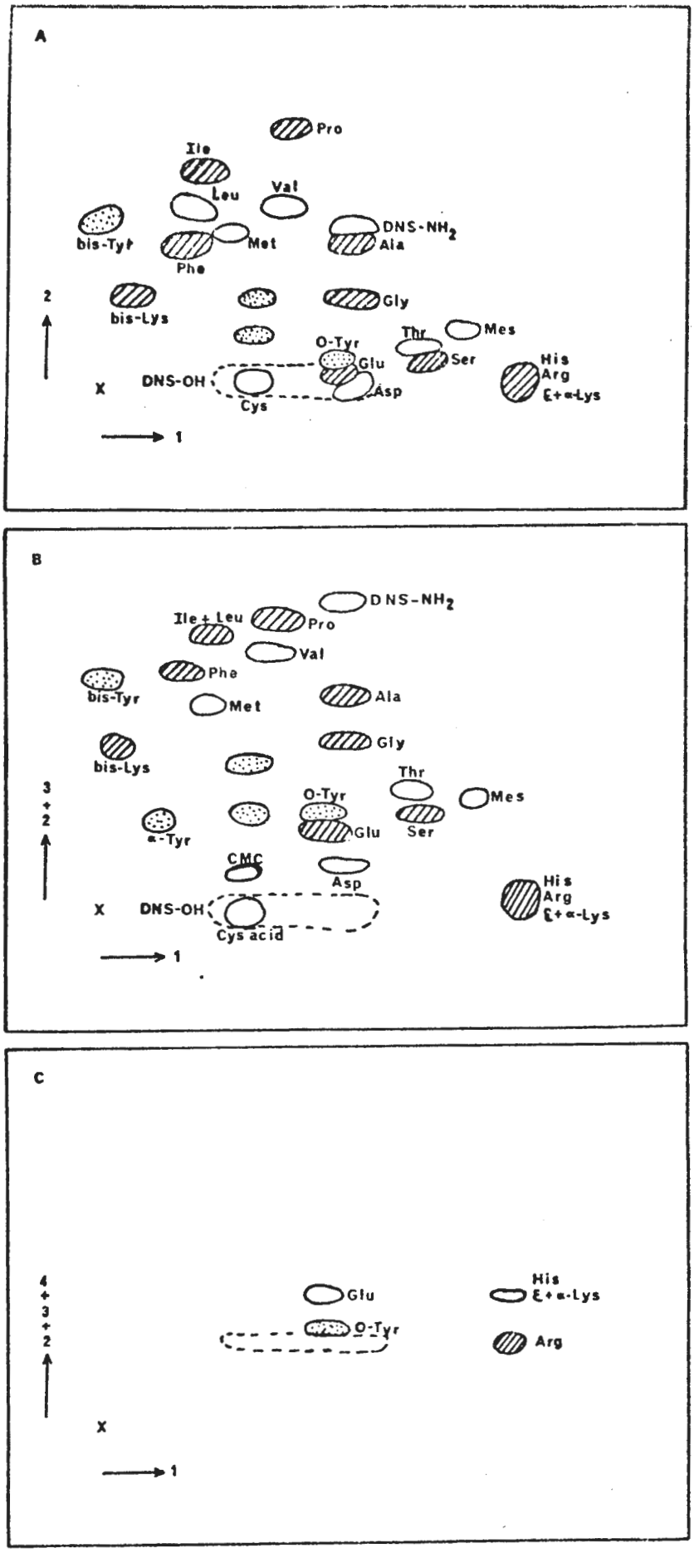


Fig. 4.1: Thin layer chromatography of standard DNS-amino acids. (For details see section 4.2.2.1.3). Cross hatched amino acids are those usually included in a standard mixture. Derivatives of tyrosine are stippled. The positions of two spots, which may be present when certain sequences of H2B histones containing tyrosine are dansylated, are also shown stippled.

DNS-arginine from ε-DNS-lysine. ε-DNS-lysine, α-DNS-lysine and DNS-histidine do not separate. Development in the 4th solvent discriminates between DNS-glutamic and O-DNS-tyrosine which can be a problem if several tyrosines are present in the sample.

A blue DNS-OH spot can be seen on all chromatograms. DNS-derivatives of tyrosine show an intense yellow-brown colour. DNS derivatives of other amino acids show a yellowish-white colour. Figure 4.1 shows chromatograms of standard DNS-amino acids after development as described above.

If valine, leucine or isoleucine are at the amino terminus of a peptide, various dipeptide DNS derivatives may be seen due to incomplete cleavage. Longer cleavage times usually will yield a decrease in the dipeptide spot and an increase in the intensity of valine, isoleucine or leucine.

At times unknown spots may have slightly different mobilities than the standards due to the presence of salts and other by-products of the DNS labelling procedure in the sample. In cases where this is a problem, such as discrimination between DNS-isoleucine/DNS-leucine and between DNS-serine/DNS-threonine,duplicate runs can be made with an appropriate standard spotted with the unknown.

4.2.2.2 Digestion with carboxypeptidase A and B

Proteins or peptides were dissolved, 100-200 nano moles/ml, in 100 mM NH_4HCO_3, pH 8. After incubation in the presence of enzymes at 37°C, aliquots of 100 µl were taken, acidified with 1 drop of glacial acetic acid and freeze dried. The free amino acids were quantitated directly on a Beckman amino acid analyzer. The results were corrected for background amino acids

after analyzing control samples taken from the carboxypeptidase- and the sample solutions prior to mixing. These control samples were incubated in parallel with the enzyme-substrate mixture.
Sources of enzyme were: carboxypeptidase A (Sigma-bovine pancreas) and B (Boehringer Mannheim - porcine pancreas).

4.2.3 Amino Acid Analysis

Hydrolyses of proteins and peptides 100-500 µg in 200 µl twice distilled constant boiling HCl containing 0,025 % phenol were performed at 110°C for 24 hours in sealed tubes (8 x 80 mm-pyrex) flushed with nitrogen and evacuated to 0,02 mm Hg. Free amino acids were identified on a Beckman amino acid analyzer using the sensitive scale (0.2). An internal standard of norleucine was used and quantitation was done by means A/D converters between the 440 and 570 mµ channels of the analyzer and a Hewlett Packard computer (Laboratory Data System 3352B).

No corrections were made for hydrolytic losses or incomplete cheavage.

4.3 FRAGMENTATION OF H2B HISTONES

4.3.1 Chemical methods

4.3.1.1 Cyanogen Bromide cleavage at methionine (Gross, 1967, Witkop, 1968)
Protein was dissolved 10 mg/ml in 70% formic acid, (w/v) a 100 fold excess (based on methionine) of CNBr (in 70% formic acid) was added and the mixture incubated with stirring, in the dark, at room temperature for 24 hours. Excess CNBr was evaporated under a stream of N_2 in a fume hood, the sample diluted 1 : 10 with distilled water, and freeze dried. Separation of peptides was accomplished on Sephadex G-100. (Fig. 2.1, 2.17).

If only small amounts of protein were available, CNBr cleavage was done in the sequencer cup after maximal sequential N-terminal degradation of the uncleaved protein. To this end the protein film remaining in the cup after a degradation cycle was treated with excess 90% acetic anhydride in pyridine (v/v) for one hour. This effectively blocked all free amino groups. The acetic anhydride-pyridine mixture was evaporated with a stream of dry nitrogen and the film finally dried in vacuo. The residue was washed with ethyl acetate and dried under nitrogen. Subsequently, 2 mg cyanogen bromide/ml 70% formic acid in water (w/v) was added in sufficient quantity to cover the protein film in the spinning cup. The cup was then covered and spun in the dark for 12 hours after which excess reagents were removed by evaporation. The film was washed with ethyl acetate, dried, and the sequencing continued at the newly created free amino group.

4.3.1.2 Dilute acid hydrolysis

To cleave peptide bonds involving aspartic acid, the protein was dissolved in 0.03 N HCl. (Schultz, 1967). To remove oxygen, the hydrolysis tube was flushed repeatedly with nitrogen and subsequently sealed and hydrolysed for 2½ hours at 110°C. The time of hydrolysis was selected after monitoring by gel electrophoresis various times of hydrolysis from 30 minutes to 8 hours. At 2½ hours no starting material was present and a pattern of cleavage was present which only became fainter with longer hydrolysis.

4.3.2 Enzymatic methods

4.3.2.1 Trypsin - digestion of maleylated material at arginine residues.

The protein or peptide was dissolved (5 mg/ml) in 100 mM $NaHCO_3$, 6 M urea at pH 8.5. A 20 fold excess of maleic anhydride (recrystallized twice from chloroform) was added slowly as a solid while the pH was maintained at about pH 8.0 with 2 M NaOH. The protein or peptide was subsequently separated from excess reagents on a Sephadex G-10 column with 100 mM NH_4HCO_3 pH 8.5 as the eluant. The protein or peptide fraction was freeze dried, then redissolved (10 mg/ml) in 100 mM NH_4HCO_3, 2 M urea pH 8.5. Trypsin (Merck No.24581 TPCK treated) was added in a 1 : 100 ratio (w/w). After 2 hours incubation at 37°C the mixture was fractionated on a Sephadex G-50 column with 100 mM NH_4HCO_3 as the eluant. (Fig. 2.2). De-maleylation of peptides (Kasper, 1975) was performed in 30% acetic acid (w/v) at 40°C for 48 hours. After freeze drying the peptides were desalted on a Sephadex G-10 column with 10 mM HCl as the eluant.

4.3.2.2 *Staphylococcus aureus* protease-digestion at glutamic acid residues.

To a protein or peptide (10 mg/ml) dissolved in 100 mM ammonium acetate pH 4.0, *Staphylococcus aureus* protease (Miles No.36-900) (Houmard and Drapeau, 1972; Drapeau, 1976) was added in an enzyme to substrate ratio of 1 : 30. After incubation for 18 hours at 37°C the mixture was freeze dried. Fractionation of peptides was on Sephadex (Fig.2.3) or on CMC (Fig. 2.13) depending on the complexity of the mixture.

4.3.2.3 Chymotrypsin-cleavage at tyrosine residues

Initially N-bromosuccinimide was used to cleave at tyrosine residues, but repeated difficulties were experienced in the identification of methionine residues in N-bromosuccinimide cleavage fragments and in sequencing beyond histidine residues. These difficulties were probably due to oxidation of methionine and to modifications of histidine. Therefore, limited chymotryptic digestion was used for the cleavage at tyrosine residues.

Typically 40 mg $H2B_{(1)}$ were dissolved in water (4 mg/ml). 0.2 mg α-chymotrypsin (1 mg/ml water) and 0.05 mg soybean trypsin inhibitor were pre-mixed for ½ hour and then added to the substrate at room temperature. The pH was kept constant at pH 8.4 by titrating with 0.01 N NaOH in a Radiometer pH-stat. The reaction was terminated after 40 minutes at which point it was estimated that an average of 5 peptide bonds per $H2B_{(1)}$ molecule had been cleaved. The reaction was terminated by lowering the pH to 2.0 followed by freeze drying. The freeze dried peptides were re-dissolved in water and freeze dried a second time. Purification of peptides was on CMC (Fig.2.9).

4.3.2.4 Thermolysin

Peptides from $H2B_{(1)}$ and $H2B_{(2)}$ *Parechinus* were dissolved in a minimum of double distilled water and 1/10th volume of 25 mM $CaCl_2$ was added. The pH was adjusted to 8.1 with 100 mM NaOH. Thermolysin (2 mg/ml) (Boehringer) was dissolved in 25 mM $CaCl_2$ and used at an enzyme to substrate ratio of 1:50 (w/w). The pH was maintained with a pH-stat. To terminate the

reaction, the pH was lowered to 2.0.

Peptides from $H2B_{(3)}$ Parechinus were cleaved as follows: Thermolysin (Serva 36015) was added to the peptide, dissolved in H_2O (4 mg/ml and the solutions adjusted to pH 8.0 with 100 mM NaOH (enzyme : substrate = 1:50 w/w). The pH was maintained at pH 8.0 for 3 hours on a pH-Stat with 100 mM NaOH. The reaction was terminated by addition of HCl until the pH reached 2.0.

4.3.3 Purification of peptides

4.3.3.1 Gel exclusion on Sephadex

Various grades of resin were packed in columns as recommended by the manufacturers. The usual eluant was 10 mM HCl pH 2.0 containing 0.001% toluene as a preservative. Maleylated material was eluted from columns with 100 mM NH_4 HCO_3 pH 8.5 to prevent de-maleylation before separation from trypsin. Protein or peptide samples were usually dissolved in a small volume of 6 M urea in 10 mM HCl and then layered carefully onto the column. The flow rate was maintained by a constant hydrostatic head (mariotte flask). Fraction size was controlled by a drop counter or by taking timed fractions. The concentration of peptides was monitored in the eluates at 206 nm and 280 nm on an LKB UVICHORD III absorptiometer or by measuring individual fractions at 230 nm.

Pooled fractions were freeze dried. Additional information about particular column runs are given in the figures in the text.

4.3.3.2 Ion exchange on CMC (Whatman CM-52)

CMC columns were equilibrated with 50 mM Na-acetate HCl pH 4.5. Various NaCl gradients were used to elute peptides. The gradient buffers were 50 mM in Na-acetate/HCl pH 4.5. Acidic or neutral peptides did not bind to the resin while the very basic CN-I peptide of $H2B_{(2)}$ Parechinus was eluted from the column with 800 mM NaCl. In general the elution order depended on net positive charge, however, those with a higher arginine content were retained more tightly.

Pooled fractions were freeze dried to reduce the volume then redissolved in a minimum volume of 10 mM HCl and desalted by gel exclusion on appropriate grades of Sephadex.

4.4 SEQUENCE ANALYSIS

4.4.1 Manual micro-Dansyl Edman degradation

Peptides were sequenced by manual subtractive Edman degradation using the method described by Chen (1976) in which the newly generated N-terminal amino acid is identified as its DNS-derivative on micropolyamide thin layer plates (4.2.2.1) (Narita et al., 1975).

4.4.2 Automatic Solid State Edman degradation

Two peptides were sequenced by solid phase techniques (Laursen, 1971). Peptide MT-I was attached via its C-terminal carboxyl group to aminopolystyrene resin (Previero et.al , 1973; Wittmann-Liebold and Lehmann, 1975). Peptide CN-I; SP-II was attached via its C-terminal homoserine lactone to the resin (Horn and Laursen, 1973).

After conversion in 20% TFA as previously described (Wittmann-Liebold et al., 1975) the PTH derivatives were identified on thin layer plates (Wittmann-Liebold et al., 1975).

Due to the purification of the PTH-amino acids in this method (Laursen 1971) by a Dowex 50 H^+ step instead of the acid extraction used by Edman and Begg (1967), the PTH derivatives of arginine and histidine cannot be identified because they remain bound to the resin.

4.4.3 Automatic (spinning cup) Edman degradation

Automatic sequential Edman degradation (Edman and Begg, 1967) was done in a Beckman model 890 sequencer as previously described (Brandt and von Holt 1974).

4.4.3.1 Operation of the sequencer

4.4.3.1.1 Purification of chemicals

The 3-dimethylamino-1-propyne was purified and the buffer prepared according to Braunitizer and Schrank (1970). Other chemicals were purified as described by Edman and Henschen (1975). Purification of chlorobutane was done as previously described (Strickland et al., 1978a).

4.4.3.1.2 Programs

Proline residues have been found to cleave more slowly than other residues during sequencing (Brandt et al., 1976). Therefore, the protein program has been modified by increasing the first acid cleavage to 190 seconds. The sequencer was also manually recycled to repeat both acid cleavage steps after every proline. Such modifications successfully eliminate overlaps due to incomplete cleavages at proline residues. The programs used are described in tables 4.1 and 4.2.

Table 4.1:

PROTEIN PROGRAM

Step	Program Statement	Step Time (sec)	Cup Speed	Step	Program Statement	Step Time (sec)	Cup Speed
1	Stop slew	2	L	40	Delay	82	H
2	Delay	6	L	41	Vac. restricted	60	H
3	Blank	2	L	42	Vac. rough	40	H
4	R4 vent.	14	L	43	Vac. fine + Fraction collector step	300	H
5	R4 press.	14	L	44	Delay	6	L
6	R1 vent.	14	H	45	R3 vent. + Fraction collector vent.	20	L
7	R1 pressurize	14	H	46	R3 press.	14	L
8	R1 deliver	6	H	47	R3 deliver	26	H
9	Blank	2	H	48	Reaction	190	H
10	Vac. restricted	30	H	49	Vac. restricted	40	L
11	Delay	6	H	50	Vac. rough	50	L
12	N_2 dry	60	H	51	Vac. fine	20	L
13	R4 deliver	28	H	52	Delay	4	L
14	Reaction	300	H	53	S3 vent.	30	L
15	R4 deliver	4	H	54	S3 press.	30	L
16	Reaction	300	L	55	S3 deliver + collect	200	H
17	R4 deliver	4	L	56	Delay	40	H
18	Reaction	300	L	57	Vac. restricted	60	H
19	R5 deliver	10	L	58	Vac. rough	60	H
20	Reaction	300	L	59	Vac. fine	60	H
21	R5 deliver	120	L	60	Delay	4	H
22	Blank	2	L	61	R3 vent	14	L
23	Vac. restricted	60	L	62	R3 press	14	L
24	Delay	4	L	63	R3 deliver	26	H
25	N_2 dry	300	L	64	Reaction	90	H
26	Vac. restricted	100	H	65	Vac. restricted	40	L
27	Vac. rough	200	H	66	Vac. rough	50	L
28	Vac. fine	100	H	67	Vac. fine	20	L
29	Blank	6	H	68	Delay	4	L
30	S1 vent.	30	H	69	S3 vent.	0	L
31	S1 press.	30	H	70	S3 press.	0	L
32	S1 deliver	200	H	71	S3 deliver + waste	200	H
33	N_2 dry	200	H	72	Delay	40	H
34	Vac. restricted	30	H	73	Vac. restricted	60	H
35	Vac. rough	140	H	74	Vac. rough	40	H
36	Delay	3	H	75	Vac. fine	200	H
37	S2 vent.	30	H	76	Vac. fine	2	L
38	S2 press.	30	H	77	Vac. fine	2	L
39	S2 deliver	440	H	78	Vac. fine	0	L

Programs for peptide and protein sequencing.

Reagent 1 = phenylisothiocyanate 5% in heptane

Reagent 3 = heptafluorobutyric acid

Reagent 4 = DMAP buffer (Braunitzer and Schrank 1970)
1 M 3-dimethylamino-1-propyne in H_2O adjusted to pH 9 with trifluoroacetic acid and subsequently diluted with n-propanol in the ratio 3:4.

Reagent 5 = N_2

Solvent 1 = benzene
2 = ethyl acetate
3 = butyl chloride

Table 4.2:

PEPTIDE PROGRAM

Step	Program Statement	Step Time (sec.)	Cup Speed	Step	Program Statement	Step Time (sec.)	Cup Speed
1	Stop slew	2	L	40	S1 deliver + collect	150	H
2	Delay	6	L	41	Delay + collect	30	H
3	Blank	2	L	42	Blank	2	H
4	R4 vent.	14	L	43	Vac. restricted	30	H
5	Delay	2	L	44	Delay	6	H
6	R4 press.	14	L	45	N_2 dry	200	H
7	R1 vent.	14	H	46	Vac. restricted	100	H
8	R1 press.	14	H	47	Vac. rough	300	H
9	R1 deliver	6	H	48	Vac. fine	300	H
10	Blank	2	H	49	Blank	2	H
11	Vac. restricted	30	H	50	Delay	6	H
12	Delay	6	H	51	S3 vent.	30	H
13	N_2 dry	60	H	52	S3 press. + Fraction collector vent and step	30	H
14	R4 deliver	28	H	53	R3 vent.	14	H
15	Reaction	300	H	54	R3 press.	14	H
16	R4 deliver	0	L	55	R3 deliver	24	H
17	Reaction	300	L	56	Reaction	80	H
18	R5 deliver	4	L	57	N_2 dry	40	L
19	Reaction	300	L	58	Vac. restricted	20	L
20	R5 deliver	10	L	59	Vac. rough	30	L
21	Reaction	300	L	60	Delay	6	L
22	R5 deliver	120	L	61	S3 deliver	25	L
23	Blank	2	L	62	N_2 dry	230	L
24	Vac. restricted	60	L	63	Vac. restricted	62	L
25	Delay	4	L	64	Vac. rough	62	L
26	N_2 dry	400	L	65	Vac. fine	62	L
27	Vac. restricted	100	H	66	Blank	2	L
28	Vac. rough	200	H	67	Delay	6	L
29	Vac. fine	600	H	68	S3 deliver + collect	150	H
30	Blank	2	H	69	Delay	30	H
31	Delay + F/C step	6	H	70	Blank	2	H
32	S1 vent	30	H	71	Vac. restricted	60	H
33	S1 press. + F/C vent	30	H	72	Delay	6	H
34	S1 deliver	20	H	73	N_2 dry	200	H
35	Blank	30	H	74	Vac. restricted	60	H
36	Blank	2	H	75	Vac. rough	300	H
37	Blank	2	H	76	Vac. fine	800	H
38	Blank	2	H	77	Vac. fine	0	L
39	Delay	6	H	78	Vac. fine	800	L

Programs for peptide and protein sequencing.

Reagent 1 = phenylisothiocyanate 5% in heptane

Reagent 3 = heptafluorobutyric acid

Reagent 4 = DMAP buffer (Braunitzer and Schrank 1970)
1 M 3-dimethylamino-1-propyne in H_2O adjusted to pH 9 with trifluoroacetic acid and subsequently diluted with n-propanol in the ratio 3:4.

Reagent 5 = N_2

Solvent 1 = benzene
2 = ethyl acetate
3 = butyl chloride

4.4.3.1.3 Modifications to peptides to improve yields

Two chemical modifications have proved useful in improving the repetitive yield when sequencing the CN-II peptides of sea urchin sperm H2B proteins (M. Strickland et al., 1977; W.N. Strickland et al., 1977). The free carboxyl groups were coupled with glycine methyl ester in the presence of 1-ethyl-3-(dimethylamino-propyl)-carbodiimide to facilitate the differentiation between the free dicarboxylic amino acids and their amides (Gibson and Anderson, 1972), at the same time the repetitive yields were improved. In addition this relatively hydrophobic peptide (CN-II) was "anchored" in the cup by modification of the ε-amino groups of lysine with 4-sulfophenylisothiocyante (Braunitzer et al., 1970).

4.4.3.2 Identification and quantitation of PTH-amino acids

To each tube in the sequenator fraction collector dithioerythritol (0.1 mg) was added to improve the recovery of PTH-serine and threonine. PTH-norleucine (100 nano moles) was also added as an internal quantitation standard. The fractions from the sequencer containing the 2-anilino-5-thiazolinone derivatives of the amino acids were dried and converted to the 3-phenyl-2-thiohydantoin amino acids by heating in 0.2 ml of 1 N HCl containing 1% (v/v) ethane-thiol under nitrogen at 80°C for 10 minutes, then extracted twice with 1 ml peroxide-free ethyl acetate. The ethyl acetate layer was dried and examined before and after silylation by gas chromatography. (4.4.3.2.1) The water layer which could contain histidine or arginine was examined by amino acid analysis. In later experiments the amino acid thiozolinones were converted to phenythiohydantoin derivatives essentially as described by Wittmann-Liebold et al.,(1975). To the dried sample, 0.2 ml of 20% trifluoroacetic (TFA) was added and after flushing with nitrogen, the mixture was incubed at 80°C for 10 minutes. The TFA

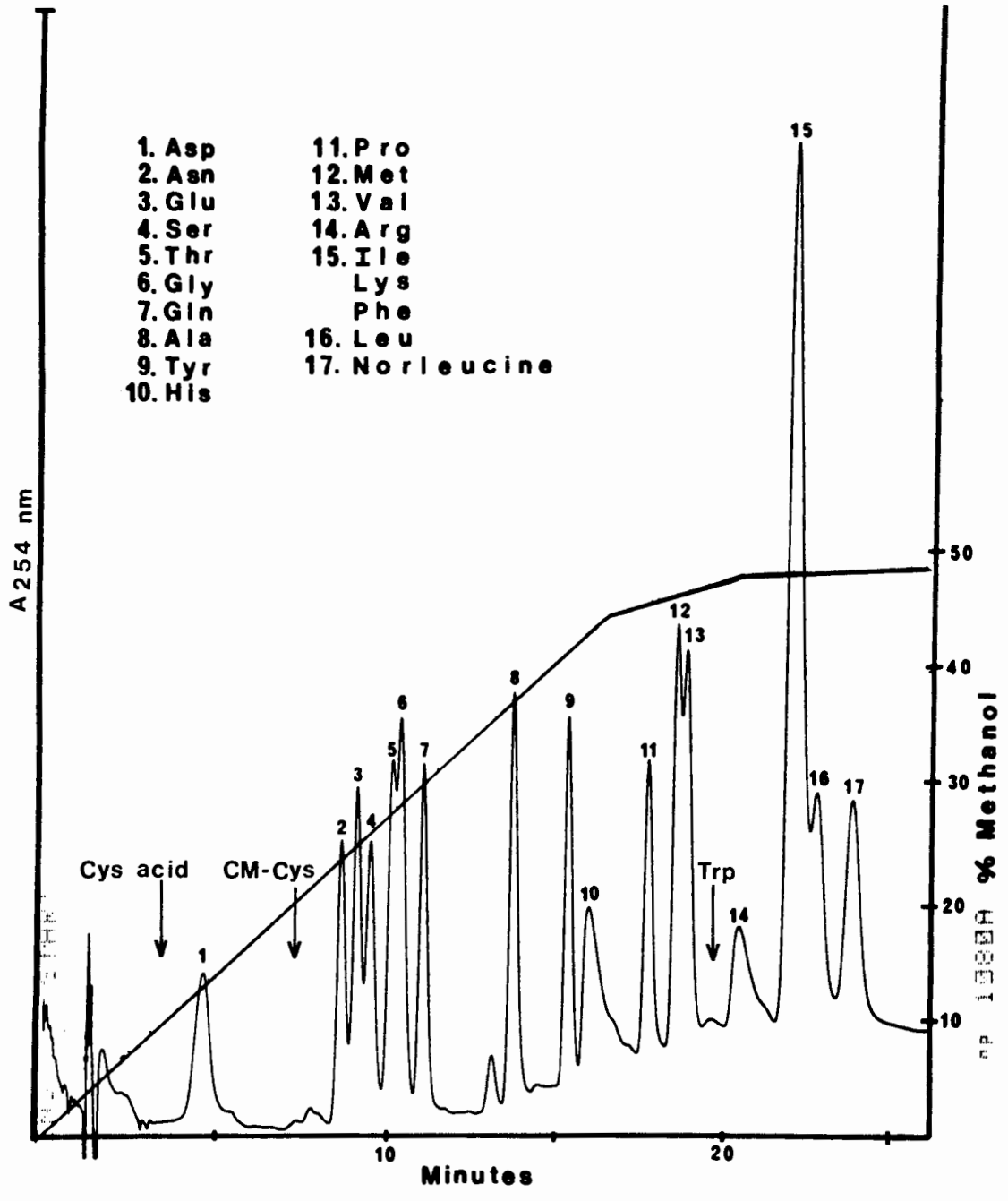


Fig 4.2:

High Pressure Liquid Chromatograph of a mixture of PTH-amino acids containing 2 nano moles of each amino acid. The elution positions of 3 other PTH-amino acid derivatives are also shown.

A Hewlett Packard Model 1084A liquid chromatograph with a Merck Li-Chrosorb RP-8 column (4.6 x 250 mm) with a particle size of 10 µm was used. Eluant A: 2.3 mM acetic acid adjusted to pH 4.75 with 1 N NaOH, eluant B: AR methanol (Merck), gradient: 5-48% B, temperature: 35°C, flow rate: 2.5 ml/minute, attenuation: 2^6 = 64 AU x 10^{-4}/cm.

was evaported under a stream of nitrogen and the dried residue was taken up in ethyl acetate for GC or in methanol when high pressure liquid chromatography (HPLC, 4.4.3.2.2) was used.

The ethyl acetate extraction step was deleted since it afforded no purification of the chemical background seen on the HPLC and had the advantage of retaining all amino acids in the same ratio with the internal standard. No chemical peaks which interfered with GC or HPLC analysis were noted. Similar results were obtained if a sample was divided and converted by each of the two methods.

4.4.3.2.1. Gas chromatography and amino acid analysis

Fractions were examined before and after silylation on a Beckman GC-45 gas chromatograph as described by Brandt & von Holt (1974) using the procedure of Pisano and Bronzert (1969). To identify arginine, histidine and lysine the unused portion of the ethyl acetate extract was pooled with the aqueous acid layer, dried and subjected to amino acid analysis after hydrolysis in constant boiling HCL-1% thioglycollic acid for 24 hours at 130^{o}C (Brandt & von Holt, 1974). Quantitation of peaks from the GC and the amino acid analyzer was by means A/D converter links between these instruments and a Hewlett Packard computer (Laboratory Data System 3352B).

4.4.3.2.2 High pressure liquid chromatography

Recently HPLC in conjunction with GC has been used. The amount of sample used for GC or HPLC varied from 2-10%. The HPLC method is similar to that described by Bridgen (1977). In a single 25 minute run 18 PTH-amino acid derivatives and the internal standard PTH-norleucine can be resolved (Fig.4.2). The PTH-derivatives in the unresolved fraction (PTH-lysine, phenylalanine, and isoleucine) are identified and quantitated by GC.

In the HPLC system used, the elution time of the acidic amino acids can be shortened by increasing the pH whereas the elution time of the basic amino acids can be delayed by decreasing the molarity of the buffer at a given pH. In a recent report on the use of methanol for HPLC (Zeeuws and Strosberg 1978) PTH-valine and methionine were not resolved. At pH 5.3 which was used in that study, these two amino acids also elute as one peak in our system. After several sequencer cycles the yields of PTH-serine and threonine usually decrease; however, in addition to the remaining amount of PTH-serine or threonine, characteristic derivatives can be seen on the HPLC. A serine derivative elutes between alanine and tyrosine. Two derivatives of threonine can usually be seen just prior to tyrosine and proline respectively. A fraction derived from sequencing chemicals elutes between glutamine and alanine.

Quantitation of PTH-amino acids was done by the internal standard method of the computer dedicated to the Hewlett Packard Model 1084 High Pressure Liquid Chromatograph.

ABBREVIATIONS

A230	=	absorbance at 230 nm
CMC	=	carboxymethyl cellulose
CNBr	=	cyanogen bromide
DMAP	=	3-dimethylamino-1-propyne
DNS) Dansyl)	=	1-dimethylaminonaphthalene-5-sulfonyl-
DTE	=	dithioerythritol
HFBA	=	heptafluorobutyric acid
PTH	=	phenylthiohydantoin
PITC	=	phenylisothiocyanate
SDS	=	sodium dodecyl sulfate
SPITC	=	4-sulfophenylisothiocyanate
TRIS	=	tris-(hydroxymethyl)-amino methane

REFERENCES

Ambler, R.C. (1972) Enzymatic Hydrolysis with Carboxypeptidases in Methods in Enzymology 25, pp 143-154.

Ando, T., Yamasaki, M. and Suzuki, K. (1973) Protamines: Isolation, Characterization, Structure and Function, Springer-Verlag, New York.

Barnes, R.D. (1974) in Invertebrate Zoology, 3rd Edition, W.B. Saunders Co. Philadelphia, London, Toronto. p.751.

Birnstiel, M.L., Schaffner, W. and Smith, H.O. (1977) DNA Sequences Coding for the H2B Histones of Psammechinus miliaris. Nature 266, 603-607.

Bloch, D.P. (1969) A Catalog of Sperm Histones. Genetics 61 Suppl 1, 93-111.

Bohm, E.L., Strickland, W.N., Strickland, M., Thwaits, B.H., van der Westhuyzen, D.R., and von Holt, C. (1973) Purification of the Five Main Calf Thymus Histone Fractions by Gel Exclusion Chromatography. FEBS Lett. 34, 217-221.

Brandt, W.F. and von Holt, C. (1974) The Determination of the Primary Structure of Histone F3 from Chicken Erythrocytes by Automatic Edman Degradation 2. Sequence analysis of Histone F3. Eur. J. Biochem. 46, 419-429.

Brandt, W.F., Strickland, W.N., Morgan, M. and von Holt, C. (1974) Comparison of the N-Terminal Amino Acid Sequences of Histone F3 from a Mammal, a Bird, a Shark, an Echinoderm, a Mollusc and a Plant. FEBS Lett. 40, 167-172.

Brandt, W.F., Edman, C., Henschen, A and von Holt, C. (1976) Abnormal Behavior of Proline in the Isothiocyanate Degradation. Hoppe-Seyler's Z. Physiol. Chem. 357, 1505-1508.

Brandt, W.F., Strickland, W.N., Strickland, M., Carlisle, L., Woods, D. and von Holt, C. (1978) A Histone Programme during the Life Cycle of the Sea Urchin. Manuscript submitted.

Brauntizer, G. and Schrank, B. (1970) Kurzmitteilurg 3-Dimethylamino-1-propin alf Puffersubftanz. Hoppe-Seyler's Z. Physiol. Chem. 351, 417.

Braunitzer, G., Schrank, B.,Ruhfus, A. (1970) Versuche Zum Vollstandigen und Automatischen Abbau von Peptiden Nach der Quadrolmethode. Hoppe-Seyler's Z. Physiol. Chem. 351, 1589-1590.

Bridgen, J. (1977) Trends in Amino Acid Sequencing. Sci. Tools LKB Instrum.J. 24, 1-6.

Butler, P.J.G. and Hartley, B.S. (1972) Maleylation of Amino Groups.in Methodsin Enzymology 25, 191-199.

Chen, R. (1976) The Sequence Determination of a Protein in a Micro Scale: The Sequence Analysis of Ribosomal protein L34 of Escherichia coli Hoppe-Seyler's Z. Physiol. Chem. 357, 873-886.

Clark, A.M. and Courtman-Stock, J. (1976) The Echinoderms of Southern Africa. British Museum (Natural History) London.

Coelingh, J.P., Monfoort, C.H., Rozijn, T.H., Leuven, J.H.G., Schiphof, R., Steyn-Parve, E.P., Braunitzer, G., Schrank, B. and Ruhfus, A. (1972) The Complete Amino Acid Sequence of the Basic Nuclear Protein of Bull Spermatozoa. Biochem. Biophys. Acta. 285 1-14.

Cohen, L.H., Newrock, K.M., Zweidler, A. (1975) Stage-Specific Switches in Histone Synthesis During Embryogenesis of the Sea Urchin. Science 190, 994-997.

Dayhoff, M.O. (1972) in Atlas of Protein Sequence and Structure (M.O. Dayhoff ed.) Biomed. Res. Found. Vol.5, Washington D.C. pp.D-278.

Dayhoff, M.O. (1976) in Atlas of Protein Sequence and Structure (M.O. Dayhoff ed.) Biomed. Res. Found. Vol.5, suppl 2, Washington D.C. pp. 1-8.

Drapeau, G.R. (1976) Protease from Staphylococcus aureus in Methods in Enzymology 45, 469-475.

Edman, P. (1950) Method for Determination of Amino Acid Sequence in Peptides. Acta Chem. Scand. 4, 283.

Edman, P. and Begg, G. (1967) A Protein Sequenator. Eur.J. Biochem 1, 80-91.

Edman, P. and Henschen, A. (1975) Sequence Determination in Protein Sequence Determination 2nd Edition (Needleman, S.B. Ed.) Springer-Verlag Berlin-Heidelberg-New York. pp.232-279.

Elgin, S.C.R., and Weintraub,H. (1975) Chromosomal Proteins and Chromatin Structure. Annual Review of Biochemistry, 725-774.

Franklin, S.G., and Zweidler, A.(1977) Non-allelic Variants of Histones 2a, 2b and 3 in Mammals. Nature 266, 273-275.

Gibson, D. and Anderson, P.J. (1972) Carboxyl Group Modification and Amide Assignments in Automated Sequencing of Proteins. Biochem. Biophys, Res. Comm. 49, 453-459.

Gray, W.F. (1972) End-Group Analysis Using Dansyl Chloride.in Methods in Enzymology 25, 121-138.

Gross, E. (1967) The Cyanogen Bromide Reaction.in Methods in Enzymology 11, 238-255.

Hnilica, L.S. (1972) in The Structure and Biological Functions of Histones CRC press, Chemical Rubber Co. pp.38-42.

Horn, M.J. and Laursen, R.A. (1973) Solid-Phase Edman Degradation: Attachment of Carboxyl-terminal Homoserine Peptides to an Insoluable Resin. FEBS Lett. 36, 285-288.

Houmard, J. and Drapeau, G.R. (1972) Staphylococcal Protease; A Proteolytic Enzyme Specific for Glutamoyl Bonds. Proc.Natl.Acad.Sci. U.S.A. 69, 3506-3509.

Hyman, L.H. (1955) The Invertebrates: Echinodermata The Coelomate Bilateria IV. McGraw-Hill. New York-Toronto-London.

Iwai, K., Ishikawa, K. and Hayashi, H. (1972) Calf Thymus Lysine and Serine Rich Histone 3. Complete Amino Acid Sequence and its Implications for Interactions with DNA. J. Biochem. 72, 357-367.

Johns, E.W. (1964) Studies on Histones 7. Preparative Methods for Histone Fractions from Calf Thymus. Biochem.J. 92, 55-59.

Johns, E.W. (1977) The Isolation and Purification of Histones in Methods in Cell Biology 16 - Academic Press, New York. pp.183-203.

Kasper, C.B. (1975) Fragmentation of Proteins for Sequence Studies and Separation of Peptide Mixtures in Protein Sequence Determination 2nd Edition (Needleman, S.B. ed.) Springer-Verlag Berlin-Heidelberg-New York. p.135.

Kistler, W.S., Geroch, M.E., and Williams-Ashman, H.G. (1973) Specific Basic Proteins from Mammalian Testes and Epididymal Spermatozoa J. Biol. Chem. 248, 4532-4543.

Kootstra, A. and Bailey, G.S. (1976) The Primary Structure of Histone H2B from Brown Trout (Salmo Trutta) Testes. FEBS Lett. 68, 76-78.

Laursen, R.A. (1971) Solid-Phase Edman Degradation on Automatic Peptide Sequencer. Eur.J. Biochem. 20, 89-102.

Matsubara, H. (1970) Purification and Assay of Thermolysin in Methods Enzymology 19, 642-651.

Nakano, M., Tobita, T., and Ando. T. (1970) Fractionation of Galline, a Protamine from Fowl Sperm and some Characterization of the Components Biochem.Biophys. Acta 207, 553-555.

Narita, K., Matsuo, H. and Nakajima, T. (1975) End Group Determination in Protein Sequence Determination, 2nd Edition (Needleman, S.B. ed.) Springer-Verlag. Berlin-Heidelberg-New York. pp. 42-55.

Panyim, S. and Chalkley, R. (1969) High Resolution Acrylamide Gel Electrophoresis of Histones. Arch. Biochem. Biophys. 130, 337-346.

Paoletti, R.A. and Huang, R.C.C. (1969) Characterization of sea urchin sperm chromatin and its basic proteins. Biochemistry 8, 1615-1625.

Platz, R.D., Meistrich, M.L. and Grimes, S.R. Jr. (1977) Low-Molecular-Weight Basic Proteins in Spermatids in Methods in Cell Biology 16- Academic Press, New York. pp.297-316.

Pisano, J.J. and Bronzert, T.J. (1969) Analysis of Amino Acid Phenylthiohydantoins by Gas Chromotography. J.Biol. Chem. 244, 5597-5607.

Previero, A., Serancourt, J., Coletti-Previero, M.A. and Laursen, R.A. (1973) Solid Phase Sequential Analysis: Specific Linking of Acidic Peptides by their Carboxyl Ends to Insoluable Resins. FEBS Lett. 33, 135-138.

Schroeder, W.A., Shelton, J.B. and Shelton, J.R. (1969) An Examination of Conditions for the Cleavage of Polypeptide Chains with Cyanogen Bromide Application to Catalase. Arch. Biochem. Biophys. 130, 551-556.

Schultz, J. (1967) Cleavage at Aspartic Acid.in Methods in Enzymology 11, 255-263.

Shires, A., Carpenter, M.P. and Chalkley, R. (1976) A Cysteine Containing H2B-like Histone Found in Mammalian Testis. J.Biol. Chem. 251, 4155-4158.

Smyth, D.G. (1967) Techniques in Enzymatic Hydrolysis in Methods in Enzymology 11, 214-236.

Stafford, H. (1960) A Guide to the Nomenclature and Classification of Organisms in Comparative Biochemistry. Vol. I. Academic Press pp. xvii - xxv.

Strickland, M., Strickland, W.N., Brandt, W.F. and von Holt, C. (1974) Sequence of the cysteine-containing portion of Histone F2a1 from the Sea Urchin Parechinus angulosus. FEBS Letters 40, 346-348.

Strickland, W.N., Strickland, M., Brandt, W.F., Morgan, M. and von Holt, C. (1974) Partial amino acid sequence of two new Arginine-Serine Rich Histones from Male Gonads of the Sea Urchin Parechinus angulosus. FEBS Lett. 40, 161-166.

Strickland, W.N., Schaller, H., Strickland, M., and von Holt, C. (1976) Partial Amino Acid Sequence of Histone H1 from Sperm of the Sea Urchin Parechinus angulosus. FEBS Lett. 66, 322-327.

Strickland, M., Strickland, W.N., Brandt, W.F., and von Holt, C. (1977) The Complete Amino-Acid Sequences of Histone $H2B_{(1)}$ from Sperm of the Sea Urchin Parechinus angulosus. Eur. J. Biochemistry 77, 263-275.

Strickland, W.N., Strickland, M., Brandt, W.F., and von Holt, C. (1977) The Complete Amino-Acid Sequence of Histone $H2B_{(2)}$ from Sperm of the Sea Urchin Parechinus angulosus. Eur. J. Biochemistry 77, 277-286.

Strickland, M., Strickland, W.N., Brandt, W.F., von Holt, C., Lehmann, A., and Wittmann-Liebold, B. (1978a)The Complete Amino-Acid Sequence of Histone $H2B_{(3)}$ from Sperm of the Sea Urchin Parechinus angulosus. Eur.J. Biochem. (in press).

Strickland, M., Strickland, W.N., Brandt, W.F. and von Holt, C. (1978b) The Partial Amino-Acid Sequences of two H2B Histones from Sperm of the Sea Urchin Psammechinus miliaris. Biochem,Biophys.Acta. (in press)

Subirana, J.A. and Palau, J. (1968) Histone-like Proteins from the Sperm of Echinoderms. Exp.Cell. Res. 53, 471-477.

Subirana, J.A. (1970) Nuclear Proteins from a Somatic and Germinal Tissue of the Echinoderm Holothuria tubulosa. Exp.Cell. Res. 63, 253-260.

Subirana, J.A., Cozcolluela, C., Palau, J., and Unzeta, M. (1973) Protamines and Other Basic Proteins from Spermatozoa of Molluscs. Biochem, Biophys. Acta 317, 364-379.

Subirana, J.A., Puigjaner, L.C., Roca, J., Llopis, R. and Suau, P. (1975) X-ray Diffraction of Nucleohistones from Spermatozoa.in The Structure and Function of Chromatin (Ciba Foundation Symposium 28) Assoc. Scientific Publishers pp.157-179.

van Helden, P., Strickland, W.N., Brandt, W.F. and von Holt, C. (1978a) Histone H2B Variants from the Erythrocytes of an Amphibian, a Reptile and a Bird. Biochem.Biophys.Acta 533, 278-281.

van Helden, P., Strickland, W.N., Brandt, W.F. and von Holt, C. (1978b) The Complete Amino-Acid Sequence of Histone H2B from the Mollusc Patella granatina. Submitted for publication.

von Holt, C., Strickland, M., Brandt, W.F. and Strickland, W.N. (1978) A New Family of Tissue Specific Histones H2B from Sea Urchin Sperm. Submitted for publication.

Witkop, B. (1968) Chemical Cleavage of Proteins. Science 162, 318-326.

Wittmann-Liebold, B. and Lehmann, A. (1975) Comparison of Various Techniques Applied to the Amino Acid Sequence Determination of Ribosomal Proteins. in Solid Phase Methods in Protein Sequence Analysis (Laursen, R.A. ed.) pp.81-90. Pierce Chemical Company Rockford U.S.A.

Wittmann-Liebold, B., Geissler, A.W., Marzinzig, E. (1975) Studies on the Primary Structure of 14 Proteins from the Large Subunit of Escherichia coli Ribosomes with an Improved Protein Sequenator and with Mass Spectrometry. J. Supramol. Struct. 3, 426-447.

Wouters-Tyrou, D., Sautière, P. and Bisèrte, G. (1976) Covalent Structure of the Sea Urchin Histone H_4. FEBS Lett. 65, 225-228.

Wouters-Tyrou, D. (1977) Les Histones De Gonades D'Oursin Psammechinus miliaris Ph.D. Thesis. University of Lille.

Zeeuws, R. and Strosberg, A.D. (1978) The Use of Methanol in High-Performance Liquid Chromatography of Phenylthiohydanloin Amino-acids. FEBS Lett. 85, 68-72.
